IntechOpen

Efficient Decision Support Systems

Practice and Challenges from Current to Future

Edited by Chiang Jao

EFFICIENT DECISION SUPPORT SYSTEMS – PRACTICE AND CHALLENGES FROM CURRENT TO FUTURE

Edited by **Chiang S. Jao**

Efficient Decision Support Systems - Practice and Challenges From Current to Future
http://dx.doi.org/10.5772/682
Edited by Chiang Jao

Contributors

Stefan Rass, Kyandoghere Kyamakya, Fadi Al Machot, Abdelrahman Osman Elfaki, Russell Yost, Tasnee Attanandana, Carol Jean Pierce Colfer, Stephen Itoga, Joerg Samietz, Antonio Jiménez, Alfonso Mateos, Jūrate Šliogeriene, Dalia Štreimikienė, Arturas Kaklauskas, Tomas Hujer, Christine Chan, Aleix Serrat-Capdevila, Juan Valdes, Hoshin Gupta, José R. Méndez, Florentino Fernández Riverola, Miguel Reboiro-Jato, Daniel Glez-Peña, Hugo López-Fernández, Daniel Jeske, Terrance P. Callanan, Li Guo, Roberto De Lotto, Emilio JosÃ© Castellano Torres, Manuel J. Barranco, Martínez Luis, Ahmad Shahi, Rodziah Binti Atan, Md. Nasir Sulaiman, Ignacy Kardel, Jaroslaw Chormanski, Dorota Miroslaw-Swiatek, Tomasz Okruszko, Mateusz Grygoruk, Martin Wassen, Alexandr Pavlovich Eremeev, Vadim Vagin, Fernando Pech-May, Crtomir Rozman, Andrej Skraba, Karmen Pažek, Miroljub Kljajić, Franc Bavec, Martina Bavec, Claire Laudy, Paweł Sitek, Miguel Gutierrez, Alfonso Duran, Burhanuddin Mohd Aboobaider, Sami M Halawani, Ab Rahman Ahmad, Urška Sancin, Bojan Dolšak, Yuxiang Huang, Ruixiang Zhu, Sheng Huang, Xiaohui Yang, Zhao Zhang, Maxime Morge, Dale Dzemydiene, Saulius Maskeliunas, Paolo Racca, Benno Kleinhenz, Thorsten Zeuner, Barbara Keil, Beate Tschöpe, Jeanette Jung

Notice

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in Croatia, 2011 by INTECH d.o.o.
eBook (PDF) Published by IN TECH d.o.o.
Place and year of publication of eBook (PDF): Rijeka, 2019.
IntechOpen is the global imprint of IN TECH d.o.o.
Printed in Croatia

Legal deposit, Croatia: National and University Library in Zagreb

Additional hard and PDF copies can be obtained from orders@intechopen.com

Efficient Decision Support Systems - Practice and Challenges From Current to Future
Edited by Chiang Jao

p. cm.

ISBN 978-953-307-326-2

eBook (PDF) ISBN 978-953-51-5534-8

Meet the editor

Dr. Chiang S. Jao, Ph.D., is chief biomedical informaticist with Tranformation Inc. based in Maryland. He has been involved in medical informatics since coming to University of Illinois at Chicago in 1992 to work on clinical decision support systems. His research was awarded the grant from National Patient Safety Foundation in investigating the matching of prescribing medications and clinical problems in electronic heath records. He was the visiting scholar in the Lister Hill National Center for Biomedical Communications, National Library of Medicine and built a standard drug-problem database based on authoritative information from approved drug package inserts. He has extensive experience as a software consultant to healthcare institutions. He is a senior member of the Institute of Electrical and Electronic Engineers (IEEE) and the American Medical Informatics Association (AMIA).

Contents

Preface

Series Preface

This series is directed to diverse managerial professionals who are leading the transformation of individual domains by using expert information and domain knowledge to drive decision support systems (DSSs). The series offers a broad range of subjects addressed in specific areas such as health care, business management, banking, agriculture, environmental improvement, natural resource and spatial management, aviation administration, and hybrid applications of information technology aimed to interdisciplinary issues.

This book series is composed of three volumes: Volume 1 consists of general concepts and methodology of DSSs; Volume 2 consists of applications of DSSs in the biomedical domain; Volume 3 consists of hybrid applications of DSSs in multidisciplinary domains. The book is shaped by the decision support strategies in the new infrastructure that assists the readers in full use of the creative technology to manipulate input data and to transform information into useful decisions for decision makers. This book series is dedicated to support professionals and series readers in the emerging field of DSS.

Preface

As innovative decision support strategies and computer technology keep continuously evolving in the 21st century, more sophisticated applications of decision support systems (DSS) are created and broadly adopted in diverse domains, to assist administrative professionals in improving decision making and problem solving skills. DSS development heavily relies on fine-tuned methodologies, appropriate semantic representations of knowledge data, and a set of well-trained expert rules in generating reliable decisions. The book seeks to explore empirical methodology developed to address current issues, to define feasible knowledge representations to undergo different data constraints, to validate a useful DSS to applicable domains after satisfactory system assessment, and to explore the practice and challenges of DSS implementation from current to predictable future.

This book is edited as a textbook so that it can be adopted in formal courses examining DSSs. It can be used by undergraduate senior and graduate students from diverse

computer-related fields. It will also be of value to established professionals as a text for self-study or for reference. This book is aimed to introduce the general concepts and methodology of developing DSS and to illustrate their use in a variety of professional fields.

Section 1, including Chapter 1 through 6, presents the concept of knowledge representations and emerging methodology to develop useful DSS. Chapter 1 aims at the issue of soft data fusion and focuses on one possible approach that allows taking into account the discrepancies observed among different pieces of data. With an overview of "soft data" characteristics in DSS, also soft data representation and fusion are outlined as several semantic representation formalisms such as semantic nets, ontologies and conceptual graphs. A case study is presented to validate the validity and usefulness of this empirical approach.

Chapter 2 deals with the issue of representing and validating the knowledge objects in a DSS in order to facilitate knowledge management tasks. Variability knowledge representation can allow the construction of decision repository with decision points and choices, and the validation of decision repository and decision making process. It provides automatic support for five operations needed in knowledge validation.

Chapter 3 presents the design of a suite of software tools that supports knowledge visualization and management. The suite of tools is able to support visualization of both static and dynamic knowledge models, and perform ontology management in a distributed framework environment. These tools address the weakness observed in existing ontological engineer support tools. A useful DSS is presented to support three-dimensional visualization of an ontology model for better ontology management.

Chapter 4 presents the fuzzy logic system from nonlinear modeling with analysis of the data to get accurate decision making. By detecting and reducing the uncertainty (for example, outliers and noisy data) by employing Fuzzy clustering algorithms, the DSS is able to validate its knowledge data with the improved quality in the performance. The weather temperature dataset was tested in assessing the proposed algorithm. The results demonstrate the improvement of better accuracy decision making in fuzzy system for weather temperature prediction.

Chapter 5 explores a fine-grained diagnosis of ontologies that assists knowledge engineers in transforming complex rules and inconsistent facts into a rule-based ontological reasoning by promoting its accuracy and reasoning performance. A generalized approach of the axiom diagnostic algorithm can reason with ontology to derive assertions from known facts and rules that control the underlying granularity and scope much easily. The examples from matrix games and reasoning games illustrate the importance of ontology and query optimization in maximizing reasoning performance.

Chapter 6 presents common-sense reasoning methods (including case-based and analogous reasoning) for diagnostics of complex object states. A model-based

diagnosis process is the comparison of predicted device behavior with its observed behavior. With the illustration of examples, the applications of reasoning algorithms for solution search are proved to apply to real-time diagnostics and forecasting.

Chapter 7 presents key aspects related with the application of case-based reasoning (CBR) methodology to the construction of adaptable decision support systems. An overview about CBR life cycle and combination strategies for constructing hybrid artificial intelligent systems and the main characteristics of successful DSSs developed following CBR principles are introduced. The practical implementation of DSS fully designed following the CBR paradigm is applied in different research areas to empower their adaptability and accuracy for solving problems in the respective fields.

Chapter 8 introduces Net Promoter Score (NPS), a remarkable metric used in a variety of industries for measuring customer advocacy. It demonstrates how a statistical classification model can be used to identify key drivers of NPS as a form of the DSS that provides analyses of NPS performance to suggest company initiatives aimed toward lifting the NPS. A case study based on a real-life data collection and analysis project is illustrated the process of building the linkage between customer satisfaction data and NPS.

Chapter 9 illustrates a declarative framework for decision support in outbound logistic supply chain management (SCM) as an example of use. It includes the concept of DSS structures for the SCM in the form of an additional layer of information. The declarative framework can be integrated with SCM and Enterprise Resource Planning as useful solutions. It proposes a very flexible and versatile means of automatic generation of decision-making models based on the contents of the database.

Chapter 10 introduces a generic model base design for the DSS aimed to the issue of maximizing revenue in the allocation of service infrastructures that offers great flexibility and allows overcoming the intrinsic dependency of the algorithm on the business process model. The hierarchical metamodel-model-instance approach proposed to define the DSS model base shows great potential and addresses the three core elements of the mentioned algorithm dependency. This approach allows defining and including in the DSS the specific parameters required in a given model without ad-hoc coding or compilation.

Chapter 11 illustrates a decision making grid (DMG) model embedded with computerized maintenance management system to aid maintenance strategies for the machines as an adoption of technology management. It demonstrates on how DMG model can be used as a decision support module in failure-based maintenance.

Chapter 12 introduces an argumentation-based mechanism for decision-making that allows concessions which is a crucial feature for negotiations. The framework is built upon assumption-based argumentation frameworks, and provides mechanisms to evaluate decisions, to suggest decisions, and to interactively explain the choice which has been made to make a certain decision, along with the concessions made to support

this choice. Argumentation provides a natural framework for hybrid systems by providing a link between qualitative objectives and its quantitative representation.

Chapter 13 presents a hybrid intelligent DSS aimed to improve the performance of academic orientation. A competency-based education model hybridizes collaborative and content-based techniques based on the textual description of subject competences. An example of the Web-based DSS is illustrated to make use of collective filtering techniques for providing better recommendations to end users with improved DSS capabilities.

Chapter 14 presents the generic multi-attribute analysis (GMAA) DSS, an additive multi-attribute utility model. GMAA can be used to solve complex decision-making problems in different areas, such economy, ecosystems and environment. The GMAA system can be used to predict the potentially optimal alternatives among the non-dominated alternatives.

This book concludes in Section 3, including Chapter 15 through 26, that introduces several interesting cases of DSS applications in diverse fields such as agriculture, environment, materials, power, etc. The information in each case is identified as a highly interdisciplinary consequence, involving the full range of human knowledge systems, beyond the scope of traditional biological, ecological, and physical disciplines. Knowledge representations and decision making process of integrated DSSs comprise several key components of resource management decision strategy. These DSSs would increase the awareness of information value for managers in effective decision making.

Chiang S. Jao
Transformation, Inc. Rockville
Maryland University of Illinois (Retired) Chicago
Illinois

Part 1

Knowledge Representations and Methodology

1

Semantic Knowledge Representations for Soft Data Fusion

Claire Laudy
Thales
France

1. Introduction

Within the fusion community, studies have long been focused on the issue of fusing data that comes from physical sensors. This type of data is also called "hard data" as the data items represent physical measurements. Recently the issue of fusing information called "soft data" has come to the fore. Soft data is generated by humans and may be expressed as natural language or semi structured data. Soft data processing is a major issue for many support systems. The new intelligence support systems for instance, aim at taking the advantage of all types of data, and among them soft data available on the World Wide Web.
The aim of this chapter is to present the issue of soft data fusion and focus on one possible approach that allows taking into account the discrepancies that may be observed among different pieces of data.
The chapter is organized as follows. Section two gives an overview of what "soft data" is, as opposed to what is commonly named "data" or "hard data" within the fusion community. Relying on existing studies, we give an overview of soft data characteristics. We also emphasise on the need that has recently been identify, to take into account soft data within decision support systems.
Section three aims at giving the status and context of soft data fusion within the wide range of fusion approaches and studies. In order to explain the context of soft data fusion first, we present some of the different models that exist that aim at classifying fusion systems. Then, we focus on studies related to the fusion of graph structures, as they appear to be key structures for soft data representation. Given the exposed context, we then describe our proposition of framework for soft data fusion which is based on three main phases: the modeling of the application domain, an association phase and finally a multi-source fusion phase.
As we will see, soft data encompasses a high level of semantic. Therefore, semantic representation formalisms are needed for soft data representation and fusion. Section four is dedicated to the description of several semantic representation formalisms such as semantic nets, ontologies and conceptual graphs.
The fifth section is dedicated to the detailed description of our proposition for soft data fusion. The three phases defined before are detailed with a proposition of approach. We first describe a case study that will be used in order to illustrate our approach. It concerns TV program description fusion, within the context of a smart TV program recommendation system. Conceptual graphs are used for domain modeling. Therefore, we will focus on this semantic representation formalism and explain how we use it. We then describe the

association phase, which is carried out relying on several similarity measures and the fusion phase which relies on an existing operation on the conceptual graphs, the *maximal join*, extended thanks to domain knowledge.
Section six describes some of the experimentations that we conducted on TV program description, in order to demonstrate the validity and usefulness of our approach.

2. Soft data

2.1 From data to information: an abstraction process

In the domain of Artificial Intelligence (AI), data, information and knowledge are central notions. Many definitions exist for the words "data", "information" and "knowledge". In this paragraph, we give our definitions of these three concepts. They are inspired from the socio-economical domain. Data, information and knowledge are points in a continuum along which an abstraction process takes place.

Abstraction is the process of highlighting some of the aspects of a thing in order to grasp its characteristics. It is somehow a process of generalization. Abstracting an observable thing leads to a general representation of this reality, which is often called a *concept*.

Data items are unprocessed and uninterpreted symbols. They are elementary descriptions of measurable properties.

Information is what data items become once they have been interpreted and contextualized so to become useful within a specific objective and for a specific user. Having information is "knowing what is happening". The information answers to questions such as "who?", "what?", "where?" and "when?"

Knowledge is a combination of information with experience and judgment. It allows reasoning among information and interpreting data in order to create new data and information items. The knowledge answers to the question "How?".

In the specific case of fusion, the notions of data, information and knowledge are also linked one to another within the process of abstraction (see Figure 1). The aim of information and data fusion is to have a representation of an external situation. This representation can be built thanks to observations of the external situation that are acquired through sensors and reported to fusion systems. Sensors are either low-level physical sensors, that report about physical measurements such as temperature or speed, or human observers that report about (some parts of) complex situations. In the first case, the physical sensors give a set of data that must be interpreted. The human sensors, on the contrary, provide interpreted information. Combining all the information items in order to deduce new information and pieces of knowledge is the role of the information fusion systems.
Both data and information fusion systems use domain knowledge in order to interpret and combine the data and information items, according to a specific aim and within a specific context. Domain knowledge is also used in order to solve inconsistencies and discrepancies among the data and information items.

2.2 Soft data: a new challenge for decision support systems

"Soft data" items are observations that are generated by humans. They may be expressed as unconstrained natural language (see Sambhoos et al. (2008)), through textual data or speech signal, but can also be made of semi constrained data items such as xml files or data bases, which are keyed in by humans through forms for instance. As soft data is provided by

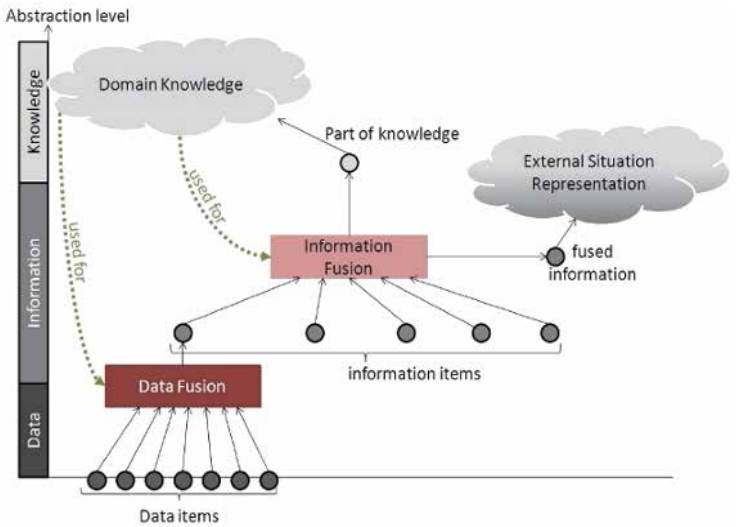


Fig. 1. Role of data, information and knowledge in the fusion process

humans, it is classified as information and knowledge according to the definitions given above.

The necessity of taking into account high-level information, which is also called "soft data" has recently been reported by the information fusion community. As stated in Pravia et al. (2008), under some circumstances, physics-based sensors are not as effective in the detection of people and complex activities for instance. In these situations, soft information sources are critical. The ability to express observed relationships is of importance for decision support systems, for inference purpose. However, most electronic sensors are feature based and do not generally provide information on relationships. Therefore, the study of soft data is of major importance.

Soft data items detections are typically qualitative, open to interpretation, and often outright inconsistent. These properties make the mathematical modeling of soft data very challenging. Studies such as Sambhoos et al. (2008), Petersen (2004) and Godbillon-Camus & Godlewski (2007) analyze the characteristics of soft and hard data, in order to clearly distinguish them. Three types of dimensions emerge from these studies:

Nature: hard information is quantitative - “numbers” (in finance these are balance sheet data, asset returns ...); soft information is qualitative - “words” (opinions, ideas, projects, comments ...); hard information is also rather “backward looking” (e.g. balance sheet data) as soft information is rather “forward looking” (e.g. business plan).

Collecting method: collection of hard data does not depend upon the context of its production, while collecting soft information includes its production context.

Cognitive factors: subjective judgment, opinions and perception are absent in hard information, whereas they are integral components of soft information.

Recent studies such as Buford et al. (2008), Sambhoos et al. (2008) and Laskey (2008) insist on the importance of such information for situation awareness and other decision support issues in general. They propose new approaches for information fusion, taking into account observations provided by humans.

3. Soft data fusion

3.1 Different levels of fusion

Many studies and systems exist that deal with data and information fusion. Each one of them focuses on a specific part of fusion. A detailed description of a large number of fusion models is proposed in Bosse et al. (2007). Within this section, we will focus on two of them. Our aim is to explain the purpose of semantic and soft data fusion, in the wide context of data and information fusion.

The North American Joint Directors of Laboratories (JDL) model for data fusion (see Hall & Llinas (2001)) is the most popular. It was first proposed in 1985 by the US Joint Directors of Laboratories Data Fusion group and revised over the years. The processing is divided into five levels described in Figure 2. The JDL model was initially proposed for the military applications but is now widely used in civil domains as well, such as business, medicine, etc.

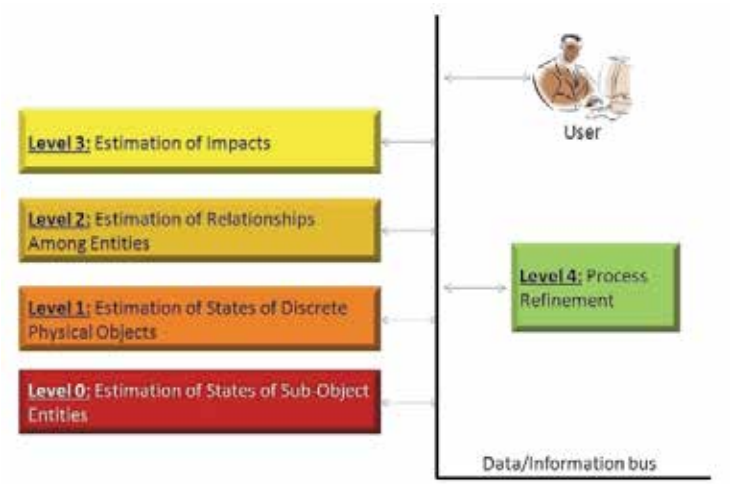


Fig. 2. The JDL data fusion process model (1999 revision)

Through its different levels, the JDL model divides the processes according to the different levels of abstraction of the data to be fused and the different problems for which data fusion is applicable. The initial levels are the following ones:

Level 0: Estimation of States of *Sub-Object Entities* (e.g. signals, features)

Level 1: Estimation of States of *Discrete Physical* Objects (e.g. vehicles, buildings)

Level 2: Estimation of *Relationships Among Entities* (e.g. aggregates, cuing, intent, acting on)

Level 3: Estimation of *Impacts* (e.g. consequences of threat activities on assets and goals)

Level 4: Process Refinement Level was initially recognized within the 1999 version of the JDL, but was then integrated to the Resource Management model levels and thus is not part of the fusion process itself.

Endsley (1995) models the data fusion process from a human perspective (i.e., Mental Model). The model has two main parts: the core Situation Awareness portion and the various factors affecting Situation Awareness. The core portion depicts three levels of mental representation: perception, comprehension and projection (see Figure 3):

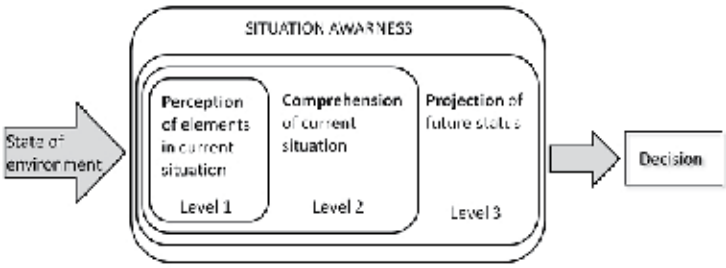


Fig. 3. Endsley's model of situation awareness (adapted from [Endsley, 1995])

Endsley's model illustrates three stages or steps of Situation Awareness formation: perception, comprehension, and projection.

Perception (Level 1): The first step in achieving Situation Awareness is to perceive the status, attributes, and dynamics of relevant elements in the environment. It also includes the classification of information into understood representations and provides the basic building blocks for comprehension and projection.

Comprehension (Level 2): Comprehension of the situation encompasses how people combine, interpret, store, and retain information. It includes the integration of multiple pieces of information and a determination of their relevance to the underlying goals. Comprehension involves a synthesis of disjointed Level 1 Situation Awareness elements through the processes of pattern recognition, interpretation, and evaluation. It includes developing a comprehensive picture of the world, or of that portion of the world of concern to the individual. Furthermore, as a dynamic process, comprehension must combine new information with already existing knowledge to produce a composite picture of the situation as it evolves.

Projection (Level 3): The third level involves the ability to project the future actions of the elements in the environment. Level 3 is achieved through knowledge of the status and dynamics of the elements and comprehension of the situation (Levels 1 and 2), and the ability to make predictions based on that knowledge.

3.2 Graphical methods for soft data fusion

Graph-based structures seam to be key structures for situation understanding and high-level information in general. Graph-based formalisms are easily readable and understandable by humans. Furthermore, graphs are a natural way to represent several ideas or objects interacting with each other. Therefore, information fusion based on graphs structures is a major stake

Sambhoos et al. (2008) relates about Inexact Graph Matching for real-time Fusion. Information items extracted from texts written in natural language are stored as RDF[1] triples. Each triple contains a subject, an object and a predicate (or relation) that exists between the subject and the object. The triples are considered as simple graphs.

Each text message may lead to the extraction of several RDF triples. These triples are then organized in more complex graph structures called "observation graphs". The larger graph is created by linking the commonly named nodes from the initial triples.

[1] Resource Description Framework (RDF) is a graph based model proposed by the World Wide Web Consortium (W3C), used to describe web resources.

Requests can then be processed on the set of observations, in order to determine whether a specific situation occurred or not. The study focuses on how to analyze the observation graph obtained after the fusion of the different information items. As for most of the studies dealing with graph homomorphism, the authors emphasize on the complexity of the underlying algorithms.
The authors propose an inexact graph matching technique using a similarity measure between the nodes and arcs. A model of a situation of interest is drawn, using a graph structure. The graph matching process then allows finding out whether this model is a sub graph of the observation graph.

3.3 Defining a high-level information fusion approach and framework

As said before, we focus on information that contains a higher-level of semantics. The approach that we proposed is optimized for information that has a high level of abstraction and that is structured. Regarding the JDL model, our approach is suitable for information fusion of levels 1 and 2 (Object Refinement and Impact Refinement). Level 2 - Comprehension - of Endsley's model for situation awareness corresponds well to our objectives as well: synthesis of perceived information items, determination of their relevance to the global objective of the user and (sub-)situation recognition through the matching with a sough-after situation.
We propose to use graphs representation formalism, and operations on graph structures. Representing information thanks to graph structures will allow us first to use existing operations and theoretical results on graphs. It will also enable to take the advantage of existing results regarding the optimization of graph algorithms. Our approach is close to the one proposed in Sambhoos et al. (2008). The aim is to fuse graphs.
However, we do not focus on the algorithmic issues of the problem, but on an other aspect: solving the conflicts that may appear in the data during the fusion process. When studying real soft data provided from operational systems, we observed that the different pieces of information, often contain conflicts. Indeed, as humans are providing the initial input data, there are very often typing mistakes or differences in the ways the same thing is reported. A simple example, is when one wants to refer to a person, a first human observer may use the person's name only, while another one will use name and surname. Therefore, the detection of conflicts among pieces of information, as well as the resolution of these conflicts within soft data fusion is of major importance.
We define hereafter the different stages that are necessary to achieve soft data fusion (Figure 4).

Situation & domain modeling is depicted by (1) and (2) on Figure 4. The situation modeling phase aims at providing a formal definition of the application domain as well as of the situations that are of interest for the user. The situations of interest are defined thanks to an empty pattern that describes their characteristics. The objective of the fusion system is to fill as much as possible this empty pattern with the observations acquired through the different sources of information.

Soft observations association is depicted by (3) on Figure 4. Observations coming from different information sources may relate to different objects. They may also relate to the same object but be reported with small differences. Therefore, it is necessary to determine whether two incoming observations rely to the same object before to fuse them.

Soft data fusion is depicted by (4) on Figure 4. When two observations are compatible and (at least partially) overlap, the multi-source information synthesis aims at building an unique view of the observed situation from them.

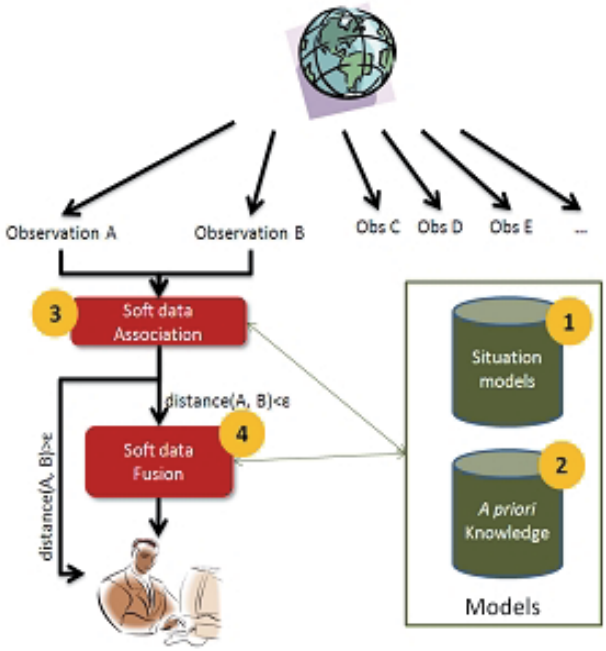


Fig. 4. General approach for situation recognition

In the remaining parts of this chapter, we emphasize on the modeling, association and fusion phases of soft observations that are not uniformly reported.

4. Domain knowledge and semantic representations

Domain knowledge has a major role within data and information fusion. Therefore, there is a need to express domain knowledge in a unique way, regardless of the different sources of information.
Furthermore, the data or information items acquired through the different sources are combined with this domain knowledge through the information process which produces new information items. This stresses the importance of having a unique formalism for knowledge representation that can also be used to represent and store the data and information that will be processed through fusion. The semantic used for representing the knowledge has to be shared between data and information as well.

4.1 Semantic networks

Within Artificial Intelligence, semantic representation formalisms were first developed in order to represent, store and automatically analyze the content of natural language.
Semantic nets (or Semantic networks) are graphical representations of interrelated concepts. A semantic network represents a taxonomy of concepts (or objects), denoted by the nodes of the network, and their properties, represented by the edges of the network. Two kinds of nodes

can be distinguished in a semantic network: concepts (or classes of objects) and individuals (or instances of the classes). The relations represented by the edges of the network are, among others, instanciation ("is a kind of") and composition ("has a").
The semantic networks were a first attempt to formalize semantic information and the reasoning process. However, as automatic processes attempted to get smarter, new semantic representations with more expressiveness and formalization were developed.

4.2 Ontologies

Ontology was initially a field of metaphysics, which aim is to study the kinds of things that exist and the relationships that can be observed among these things. Applied to the field of computer science, an ontology is a structured set of concepts that model a specific domain. Since the mid 70's, AI research scientists have had the need to capture knowledge, in order to provide "intelligent" capabilities to computers. Studies were achieved that aim at providing the ability to store general and domain specific knowledge, in a way that is understandable both by humans and computers.
An ontology captures the model of a specific domain as a catalog of categories of entities and the semantics associated with this domain. This allows making inferences and reasoning about the domain. The main components of an ontology are:

- individual entities,
- classes, which are sets or collections of entities
- attributes, which are properties of the entities represented by the different classes
- relations, which are relationships among different classes and
- events that change the state of some of the properties of entities or that modify relationships among entities.

Within the fusion community, works such as Matheus et al. (2003) insist on the importance of using ontologies to represent knowledge. For the military domain, several ontologies were developed, such as the JC3IEDM2 ontology (MIP (2005)) and the "SAW Ontology" described in Matheus et al. (2003).

4.3 Conceptual graphs

The conceptual graphs formalism is a model that encompasses a basic ontology (called *vocabulary*), graphs structures and operations on the graphs. The vocabulary defines the different types of concepts and relations that exist in the modeled application domain, while the graphs allow representing observations.
The conceptual graphs where proposed by John Sowa in Sowa (1976) as a graphical representation of logic, which was inspired by Peirce Peirce (1932). They allow representing knowledge in a easily readable manner for humans, experts of specific application domain, but non experts of knowledge representation formalisms. In our work, we use the conceptual graphs. The numerous studies achieved regarding graph algorithms and conceptual graphs in particular (Chein & Mugnier (2008)), lead us to use this model.

5. Using semantic representations for soft data fusion

5.1 Case study

We applied our proposition for soft data fusion within a TV program recommendation system. While the number of channels that one can access increases very fast, the aim of the system

is to help the users in choosing the programs that they would enjoy seeing. Based on background information and the description of a new program, the system evaluates whether the new TV program is of interest to a specific user. The recommendation system can be coupled with a Personal Video Recording system, so that interesting programs are recorded even when the users are not looking at TV.

The description on which the interest of a program is evaluated must therefore be very detailed concerning the content of the program itself. It should also be as precise as possible concerning the broadcast times so that the recording system can indicate the right slots of time.

Within the recommendation system, we used two sources of information. The first one, called DVB stream, is the live stream of metadata associated with the video stream on the TNT (Télévision Numérique Terrestre). Figure 5 shows an example of the information available on the DVB stream. The DVB stream gives descriptions of TV programs containing schedule and title information. It is very precise concerning the begin and end times of programs and delivers information about the technical characteristics of the audio and video streams. However, no description of the content of the program is given.

For each TV channel, it gives the descriptions of the currently playing program as well as the following one. The information on this source is constantly being updated. In particular, the scheduling times of the subsequent programs are adjusted.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<tv generator-info-name="TSReader">
[...]
  <channel id="1537-TF1">
    <display-name lang="en">TF1</display-name>
    <transport-stream-ID>6</transport-stream-ID>
    <signal-info>-0.0 MHz</signal-info>
  </channel>
  <programme start="20061127063959" stop="20061127064753" channel="1537-TF1">
    <title>Jt matin</title>
    <desc> |-0.0 MHz</desc>
  </programme>
  <programme start="20061127064754" stop="20061127083027" channel="1537-TF1">
    <title>TF I JEUNESSE</title>
    <desc>Au sommaire «Franklin». «Tabaluga». «Dora». «Bob l'éponge».|-0.0 MHz</desc>
  </programme>
[...]
</tv>
```

Fig. 5. Example of an initial observation on TNT metadata

The second source of information is an online TV magazine. The descriptions contain information about the scheduling of the programs, their titles and the channels on which they are scheduled. They also contain details about the contents (summary of the program, category, actors, presenters etc). This source describes all the TV programs scheduled on all the TV channels during one week starting from the current day. The TV program descriptions may be updated once a day.

On the XML initial observations, we can see the information that we are going to fuse. For instance, the beginning time of the TV program appears inside the <programme> marker, as "start" attribute, the title is between the <title> markers, ...

5.2 Domain modeling / Representing knowledge with CG

In this section, we propose to use the Conceptual Graphs model in order to achieve the step preliminary to situation recognition: *domain and situation modeling*. We briefly describe the model, using the formalization proposed in Chein & Mugnier (1992) and Chein & Mugnier (2008). As said before, the conceptual graphs model was initially proposed in order to provide a logical system able to represent and process natural language. Therefore, it is particularly well adapted to the representation and processing of soft data.
The conceptual graphs model is essentially composed of an ontology called the *vocabulary* hereafter and the graphs themselves, containing concepts and relation nodes. We detail hereafter these general notions and their notations.

Fig. 6. Example of conceptual graph

5.2.1 Concepts, relations and vocabulary

The term "concept" is used to refer to a concept node. The concepts represent the "things" or entities that exist. A concept is labeled with two components: the concept's type and the individual marker. The conceptual type defines the category to which the entity belongs. For instance, in Figure 6 the concept [Title: "TF! Jeunesse"] is an instance of the category Title. Its concept type is "Title". The individual marker relates a concept to a specific object of the world. The object represented by [Title: "TF! Jeunesse"] has the name (or value) *"TF! Jeunesse"*. The first order logic form of the concept is $\text{Title}(\text{"}TF!Jeunesse\text{"})$.
The individual markers may also be undefined. An undefined or generic individual marker is either blank or noted with a star *. It represents the existential quantifier. For instance, [Title] or [Title : *] stands for the following first order logic expression $\exists x, \text{Title}(x)$

The term "relation" is used to refer to a relation node. The relation nodes of a conceptual graph indicate the relationships that hold between the different entities of the situation that is represented. Each relation node is labeled with a relation type that points out the kind of relationship that is represented.
In this work, we consider binary relations. The arcs that link relations to concepts nodes are arrows, allowing distinguishing the source and target concept nodes.

The notion of vocabulary was defined in Chein & Mugnier (2008), as an extension of the *support* introduced in Chein & Mugnier (1992), which was itself based on Sowa's semantic network (Sowa (1984)). The concept types and the conceptual relation types, which are used to label the concept and relation nodes are organized in hierarchies. We restrict our approach to relation types that are unordered. Therefore, we manage only one hierarchy that contains the concept types.
The partial order that holds among the set of conceptual types is interpreted as a relation of specialization: $t_1 \leq t_2$ means that t_1 is a specialization of t_2, that is to say that any instance of the class denoted by t_1 is also an instance of the class denoted by t_2.

The ordered set of concept types is denoted T_C, the set of relation types is denoted T_R and the set of individual makers that are used to labeled the concept nodes is denoted $\mathcal{I}$.

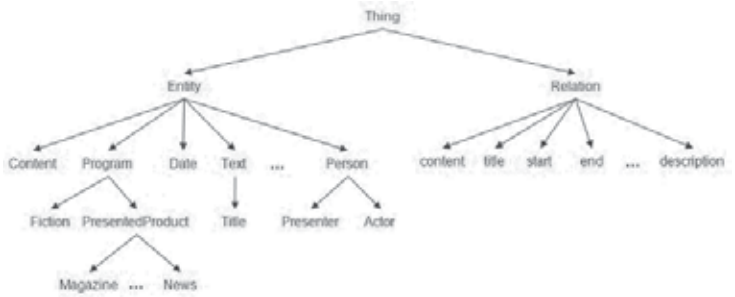


Fig. 7. Concept type hierarchy for TV programs

Figure 7 depicts an example of the hierarchy that contains the concept types used to describe TV programs.

5.2.2 Basic conceptual graphs

Several families of conceptual graphs (CG) exist, that allow describing different kinds of knowledge. Whithin this work, we focus on the basic graphs (defined in Chein & Mugnier (2008)). Basic conceptual graphs are bipartite graphs containing concept and relation nodes. Figure 6 gives an example of a basic graph. The rectangular boxes represent concept nodes and the ovals represent relation nodes.

Conceptual graphs allow to express logical formulas. Any conceptual graph can thus be translated into logic. The graph in Figure 6 for instance can be expressed in First Order Logic as follows:

$$\exists x, \exists y(\text{Content}(x) \wedge \text{Title}(\text{"}TF!Jeunesse\text{"})$$
$$\wedge\, \text{title}(x, \text{"}TF!Jeunesse\text{"}) \wedge \text{Text}(y) \wedge \text{description}(x, y))$$

A basic conceptual graph G is defined by a 4-uple over a vocabulary $V = (T_C, T_R, \mathcal{I})$: $G = (C_G, R_G, E_G, l_G)$ and is such that:

- (C_G, R_G, E_G) is a finite undirected and bipatite multigraph. C_G is the set of concept nodes. R_G is the set of relation nodes, and E_G is the set of edges.
- l_G is a naming function of the nodes and edges of the graph G which satisfies:
  1. A concept node c is labeled with a pair $(type(c), marker(c))$, where $type(c) \in T_C$ and $marker(c) \in \mathcal{I} \cup \{*\}$.
  2. A relation node r is labeled by $l(r) \in T_R$. $l(r)$ is also called the type of r.
  3. The degree of a relation node r is equal to the arity of r.
  4. Edges incident to a relation node r are totally ordered and labelled from 1 to the arity of r.

Given the order on T_C, the concept nodes that can be defined on a $T_C \times \{\mathcal{I} \cup \{*\}\}$ are partially ordered by the generality relationship. For instance, as the the concept type `Text` is greater (i.e. more general) than `Title` (see Figure 7) and the generic marker * is greater than any individual marker of $\mathcal{I}$, we have for instance:
`[Text: *]` $\geq$ `[Text : "News"]` $\geq$ `[Title : "News"]`, but `[Text : "News"]` and `[Title: *]` are not comparable.

5.3 Soft data association

In this section, we focus on the *soft observation association* phase described earlier. The aim is to compare and analyze the different observations and decide which ones should be fused. More precisely, we detail here the comparison of two observations, taking into account the domain knowledge previously modeled as well as the values of the elements (i.e. nodes and edges) that make them up. This allows us checking their fusability.

We first describe here our proposition for a similarity measure between two conceptual graphs. Then we show how to use this measure in order to test the compatibility of two graphs in the association phase.

All the measures that we propose within this section are normalized. Extended proofs of this property are available in Laudy (2010)

5.3.1 Comparison of concepts

To measure the similarity between two concepts, we propose to compare their conceptual types, their values as well as their immediate neighborhood. The study of the neighborhood gives clue about the context in which a concept is used.

5.3.1.1 Comparison of conceptual types: $diss_{type}$

We first describe how to compare two concepts, regarding their difference, through dissimilarity processing. The dissimilarity between conceptual types is used to measure how much two situations are different. We adapt the distance between types proposed by Gandon et al. (2008), in order to obtain a normalized dissimilarity measure.

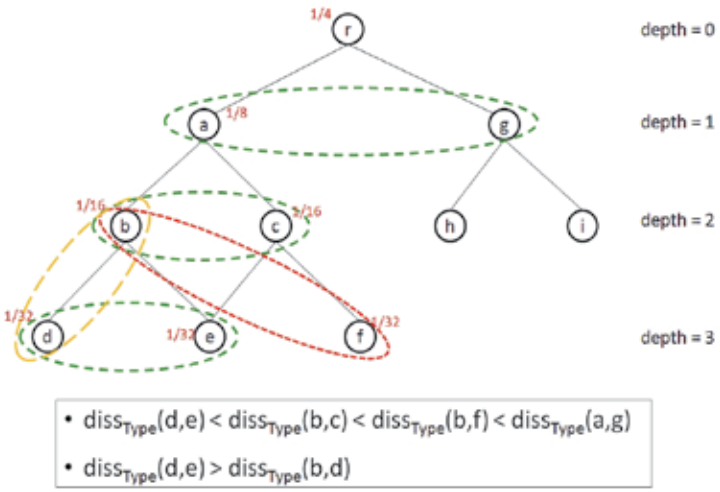


Fig. 8. Constraints on the dissimilarity over conceptual types

The main idea, is that the difference between two concepts is processed according to the number of edges that separate them from their nearest common parent. Furthermore, the deepest this common parent is in the lattice of types, the smallest the difference is between the two compared types.

The difference between two types with a nearest common parent of a depth `d` in the type lattice is always smaller than the difference between two types with a nearest parent of depth of `d-1`, whatever the number of edges between the types and their parents is.

As an illustration, looking at the figure 8, we want to have the following inequalities:

- $\text{diss}_{\text{type}}(d,e) < \text{diss}_{\text{type}}(b,c) < \text{diss}_{\text{type}}(b,f) < \text{diss}_{\text{type}}(a,g)$
- $\text{diss}_{\text{type}}(d,e) > \text{diss}_{\text{type}}(b,d)$

Definition 5.1. *The dissimilarity* $\text{diss}_{\text{type}} : T_C^2 \rightarrow [0,1]$*, where* T_C *is a set of conceptual types is defined as follows:*
$\forall (t_1,t_2) \in T_C \times T_C$,

$$\text{diss}_{\text{type}}(t_1,t_2) = \sum_{t_i \in \langle t,t_1 \rangle, t_i \neq t} 2^{-2-\text{depth}(t_i)} + \sum_{t_i \in \langle t,t_2 \rangle, t_i \neq t} 2^{-2-\text{depth}(t_i)}$$

with

- $t \in T_C$ *the nearest common parent of* t_1 *and* t_2
- $\langle t,t' \rangle$ *is the shortest path between* t *and* t'
- $\text{depth}(t_i)$ *is the depth of* t_i *in the type hierarchy, with* $\text{depth}(Entity) = 0$.

5.3.1.2 Similarity between two referents

The similarity between the values of two concepts depends on the application domain and the data type used to express the individual markers. Therefore, several similarity measures between referents are defined.
If, at least, one of the referents of the concepts is undefined, the value of the similarity is equal to 1.

Definition 5.2. $\forall (c_1,c_2) \in \mathcal{C}^2$, $c_1 = [t_1 : v_1]$, $c_2 = [t_2 : v_2]$ *and* $(t_1,t_2) \in T_C \times T_C$,

$$\text{sim}_{\text{ref}}(v_1,v_2) \begin{cases} = 1 \textit{ if } v_1 \textit{ or } v_2 \textit{ is undefined.} \\ = \text{sim}_{\text{ref}strings}(v_1,v_2), \\ \textit{or } \text{sim}_{\text{ref}num}(v_1,v_2), \\ \textit{otherwise.} \end{cases}$$

With $\text{sim}_{\text{ref}strings}$ *and* $\text{sim}_{\text{ref}num}$ *two similarity measures for referents described hereafter.*

In the next sections, we define only the measures used within the case study. For a detailed definition and description of the other measures, see Laudy (2010).
Similarity of "String" referents
The idea of $\text{sim}_{\text{ref}string}$ is to say that, if one of the strings contains a large part of the other one, the two strings should be declared sufficiently similar in order to be fused. The measure relies on the proportion of substrings shared between the two referents, regarding the length of the smallest one.

Definition 5.3. $\text{sim}_{\text{ref}strings} : \mathbb{S}^2 \rightarrow [0,1]$ *is defined as follows:*
$\forall (s_1,s_2) \in \mathbb{S}^2$, *where* $\mathbb{S}$ *is the set of all strings,*

$$\text{sim}_{\text{ref}strings}(s_1,s_2) = \frac{\text{lengthComSubString}(s_1,s_2)}{\min(\text{length}(s_1),\text{length}(s_2))}$$

where

- min *is such that*
  $\min : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$
  with $\forall (x, y) \in (\mathbb{N} \times \mathbb{N})$ *such that* $\min(x, y) = y$ *if and only if* $x \geq y$ *and* $\min(x, y) = x$ *in other cases.*
- length *is defined as follows:*
  $\text{length} : \mathbb{S} \to \mathbb{N}$ *such that* $\forall s \in \mathbb{S}, \text{length}(s) = x$*, with x the number of characters in s and where* $\mathbb{S}$ *is the set of all possible strings.*
- lengthComSubString *is defined as follows:*
  lengthComSubString :
  $$\begin{aligned} \mathbb{S}^2 &\to \mathbb{N} \\ (s, s') &\to \textstyle\sum_i \text{length}(s_i^*) \end{aligned}$$
  with $s_i^* \in S$ *such that the four following conditions are fulfill:*
  1. s_i^* *is a substring of both s and* s'
  2. s_i^* *contains at least two characters*
  3. s_i^* *is maximal (i.e. there is no other string that fulfill the conditions and is longer)*
  4. *the order in which the substrings appear in the two strings is preserved*

As an illustration of this second measure, let us consider two titles. Their similarity is the following one.

$$\begin{aligned} &\text{sim}_{\text{ref}_{strings}}(\text{“}The\ news\text{''}, \text{“}News\text{''}) \\ &= \frac{\text{lengthComSubString}(\text{“}The\ news\text{''},\text{“}News\text{''})}{\text{length}(\min(\text{“}The\ news\text{''},\text{“}News\text{''}))} \\ &= \frac{4}{4} \\ &= 1 \end{aligned}$$

Similarity of dates and numerical referents

To compare the numerical referents and dates given as numerical values, we rely on a distance and a threshold, that represents the tolerance with which two values may differ atmost.

Definition 5.4. *Let t be the threshold defined by the end user.* $t \in \mathbb{R}^{+*}$. $(v_1, v_2) \in \mathbb{R}^2$ *are two numerical values that must be compared.*
The function $\text{sim}_{\text{ref}_{num}} : \mathbb{R}^2 \to [0, 1]$ *is defined as follows:*

$$\text{sim}_{\text{ref}_{num}}(v_1, v_2) = \begin{cases} 0 \text{ if } |v_1 - v_2| \geq t \\ 1 - \frac{|v_1 - v_2|}{t} \text{ otherwise} \end{cases}$$

5.3.1.3 Similarity regarding the context of the concepts

In order to compare the context in which the two concepts are expressed, we propose to compare their immediate neighborhood. Intuitively, the similarity measure of two concepts regarding their context is processed by measuring the proportion of relations linked to the concepts and that have the same type and the proportion of relations that have different types.

Definition 5.5. *The similarity of a node* c_1 *of the graph* G_1 *and the node* c_2 *of the graph* G_2*, regarding their neighborhood is given by the function* $\text{sim}_{\text{context}} : \mathcal{C}^2 \to [0, 1]$ *defined as follows.*
Let R_1 *(respectively* R_2*) be the set of relations neighboring the concept node* c_1 *(respectively* c_2*). We define* $R_1^{\varnothing}$ *(respectively* $R_2^{\varnothing}$*), the union of the set* R_1 *(resp.* R_2*) and set containing the empty element noted* $\varnothing$*.*
Let $\mathcal{R}$ *be a symmetric relation between the elements of* $R_1^{\varnothing}$ *and* $R_2^{\varnothing}$ *such that*

- *The types of the relations are equals:* $x\mathcal{R}y \Leftrightarrow l(x) = l(y)$
- *one element of* R_1 *is related to at most one element of* R_2*:* $\forall x \in R_1, (\exists! y \in R_2$*, such that* $x\mathcal{R}y) \vee x\mathcal{R}\varnothing$
- *one element of* R_2 *is related to at most one element of* R_1*:* $\forall y \in R_2, (\exists! x \in R_1$*, such that* $x\mathcal{R}y) \vee \varnothing\mathcal{R}y$

To define the similarity measure regarding the context of two concepts, we use the two sets INTERSEC *and* COMPL*, defined as follows as follows.*
INTERSEC *is a set of couples of relations nodes of* R_1 *and* R_2 *that are related through the* $\mathcal{R}$ *relation.*
Let INTERSEC $= \{(x,y) \in R_1 \times R_2 | x\mathcal{R}y$ *with* $\forall (x,y) \in R_1 \times R_2, (x',y') \in R_1 \times R_2$*, if* $x\mathcal{R}y \wedge x'\mathcal{R}y', x = x' \Leftrightarrow y = y'\}$.
COMPL *is the set of relations that could not be related through* $\mathcal{R}$
Let COMPL $= \{(x,y) \in R_1^{\varnothing} \times R_2^{\varnothing} | \nexists y' \in R_2$ *such that* $(x,y') \in$ INTERSEC $\wedge \nexists x' \in R_1$ *such that* $(x,y') \in$ INTERSEC $\wedge (x = \varnothing \oplus y = \varnothing)\}$
The similarity of the context of c_1 *and* c_2 *is then defined according to the cardinality of the two sets* INTERSEC *and* COMPL*:*

$$\mathrm{sim}_{\mathrm{context}}(c_1, c_2) = \frac{|\mathrm{INTERSEC}|}{|\mathrm{INTERSEC}| + |\mathrm{COMPL}|}$$

5.3.1.4 Similarity of concepts

To compare two concepts, now, we use a similarity measure that combines all the measures described above.
The order of importance of the component of two concepts, when processing their similarity is the following one:

1. their concept types
2. their referents
3. the context in which they are used

To account for this hierarchy of importance, within the similarity measure $\mathrm{sim}|_{\mathrm{gene}}$, we apply different coefficients to the individual similarity (and dissimilarity) measures: a coefficient of 4 is applied to the part accounting for the similarity of the concept types, 2 to the part accounting for the referents and 1 for the contexts. In order to keep a normalized similarity, the the similarity score processed as described above is divided by 7.

Definition 5.6. *The similarity measure* $\mathrm{sim}|_{\mathrm{gene}} : \mathcal{C}^2 \rightarrow [0,1]$*, where* $\mathcal{C}$ *is a set of concepts defined on the same vocabulary, is expressed as follows:*
$\forall (c_1, c_2) \in \mathcal{C}^2$ *such that* $c_1 = [t_1 : v_1]$ *and* $c_2 = [t_2 : v_2]$*,*

- *If the most specific common parent of* t_1 *and* t_2 *is the root of the type hierarchy, we have* $\mathrm{sim}|_{\mathrm{gene}}(c_1, c_2) = 0$.
- *Otherwise, we have* $\mathrm{sim}|_{\mathrm{gene}}(c_1, c_2) = \frac{4(1-\mathrm{diss}_{\mathrm{type}}(t_1,t_2))+2*\mathrm{sim}_{\mathrm{ref}}(v_1,v_2)+\mathrm{sim}_{\mathrm{context}}(c_1,c_2)}{7}$

where $\mathrm{diss}_{\mathrm{type}}$, $\mathrm{sim}_{\mathrm{ref}}$ *and* $\mathrm{sim}_{\mathrm{context}}$ *are the similarity and dissimilarity measures defined above.*

5.3.2 Graph association

On figure 9, we can see an example of the need for the association phase. Graphs $g1$ and $g2$ represent two TV program descriptions that we attempted to fuse. The result of the fusion is given by the graph $g3$, which depict a badly formed TV program. Indeed, this fused TV program has two begin and two end dates. Furthermore, looking at these two TV program descriptions, it is obvious that they are not compatible and should not be fused because they describe two different TV programs. Our aim is to provide a method that will enable discriminating the observations that can be fused from the ones that are obviously not compatible thanks to the *association* phase.

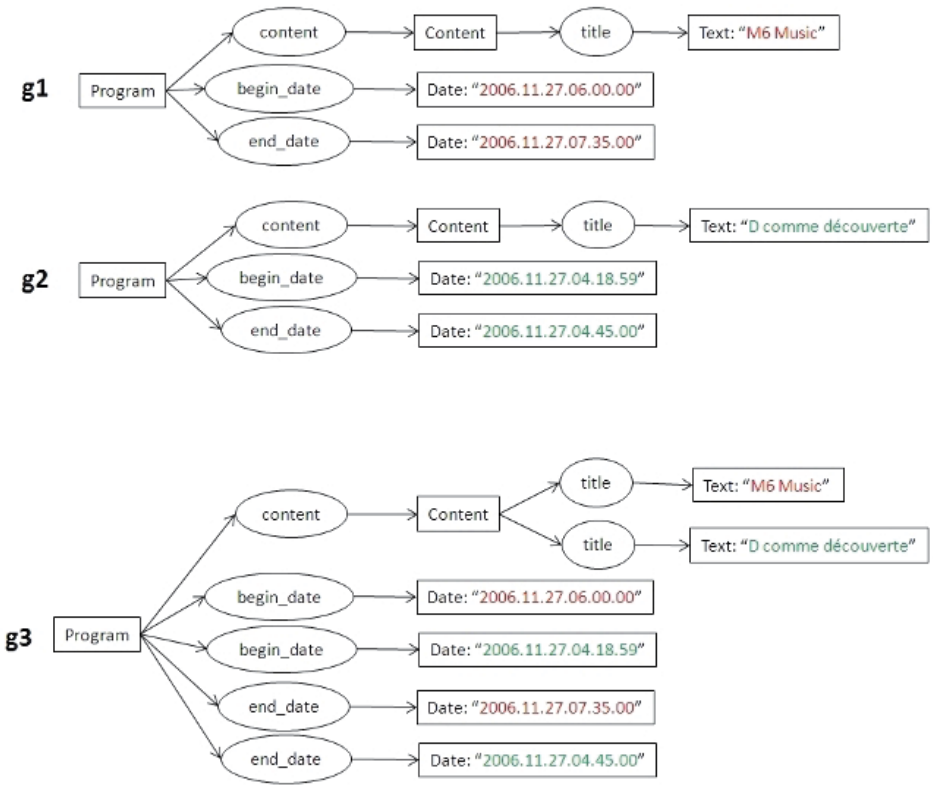


Fig. 9. Incompatible graphs

Several similarity measures between general graphs and conceptual graphs have been proposed (for instance in Sorlin et al. (2003), Gandon et al. (2008) and de Chalendar et al. (2000)). Through our proposition, we focus on the local similarity of the different pairs of nodes. We propose to compute the similarity between two graphs with regards to the best matching of their nodes. Intuitively, we process the similarity of two graphs by maximizing the similarity of their couples of concepts.

Definition 5.7. $\text{sim}_{\text{Graph}} : \mathcal{G}^2 \rightarrow [0,1]$, *where $\mathcal{G}$ is a set of graphs defined on the same vocabulary, is the function defined as follows:*
Let G_1 and G_2 be two graphs to compare. C_1 (resp. C_2) is the set of concepts of G_1 (resp. G_2) and $|C_1|$ (resp. $|C_2|$) is the number of concepts in the graph G_1 (resp. G_2).
We rename G_1 and G_2 into G_1' and G_2' such that

- *if* $|C_1| \leq |C_2|$, $G'_1 = G_1$ *and* $G'_2 = G_2$
- *else* $G'_1 = G_2$ *and* $G'_2 = G_1$

C'_1 *(resp.* C'_2*) is the set of concepts of* G'_1 *(resp.* G'_2*).*
$\mathrm{sim}_{\mathrm{Graph}}(G_1, G_2) =$

$$\frac{\sum_{c_1 \in C'_1} (\max_{c_2 \in C'_2} p(t_{1,2}) * \mathrm{sim}|_{\mathrm{gene}}(c_1, c_2))}{\sum_{\substack{(c_1,c_2) \in C'_1 \times C'_2 | \\ \forall c_3 \in C'_2, c_3 \neq c_2 \wedge \\ \mathrm{sim}|_{\mathrm{gene}}(c_1,c_3) < \mathrm{sim}|_{\mathrm{gene}}(c_1,c_2)}} p(t_{1,2})}$$

where $p(t_{1,2})$ *is the weight associated with the conceptual type* $t_{1,2}$ *and that allows giving more or less importance to some of the concepts, according to the application domain;*

As a matter of example, let us consider the two graphs $g1$ and $g2$ depicted on figure 10. To compute their similarity, we process the similarities of the different possible pairs of concepts. The table on figure 11 shows the similarity scores, using $\mathrm{sim}|_{\mathrm{gene}}$, of all the pairs of concepts that can be matched between the two graphs.

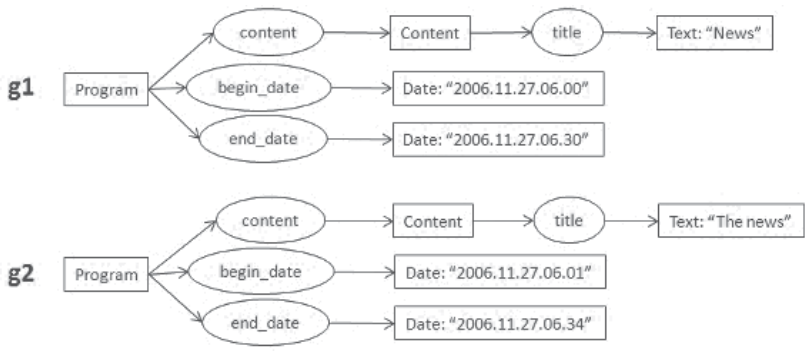


Fig. 10. Processing similarity - input graphs

The matching surrounded in red continuous line has a similarity score of 0,92 and is the best that can be found.

5.4 Multi-source synthesis

The *multi-source information synthesis* phase relies on a fusion process that is an extension of the *maximal join* operation initially described by John Sowa (Sowa (1984)). The structures and contents of the two graphs to be fused are compared relying on homomorphism search. Redundant parts are fused (i.e. added only once) into the resulting fused graph and complementary elements are all added.

5.4.1 Projection based operations on conceptual graphs and maximal join

To fuse two graphs, we first have to find all the couples of nodes of the two graphs that represent the same parts of the TV program description. Doing so, one should look, not only at the identical nodes, but also at the ones that represent the same thing, potentially with different levels of precision. For instance, in Figure 12 the `[Program]` and `[Entity: P1]` concepts represent the same object of the world, (a TV program which referent is `P1`).

Sim(g1, g2)	[Program]	[Content]	[Text = "News"]	[Date = 2006.11.27.06.00]	[Date = 2006.11.27.06.30]
[Program]	1	0	0	0	0
[Content]	0	1	0	0	0
[Text = "The news"]	0	0	[4 + 2*(4/4) +1] /7 = 1	0	0
[Date = 2006.11.27.06.01]	0	0	0	[4+ 2*(1-(1/5))+1]/7 = 0,9	0
[Date = 2006.11.27.06.34]	0	0	0	0	[4+2*(1-(4/5))+1]/7 = 0,7

Fig. 11. Processing similarity - results

Matching the two graphs, according to these couples of nodes, should also keep the structure of the graphs. Arcs between nodes should not be deleted or modified. For instance, given that (`[Program]`, `[Entity: P1]`) and (`[Content]`, `[Content]`) are two couples of nodes that are compatible, the edge between `[Program]` and `[Content]` must have an equivalent between `[Entity: P1]` and `[Content]`, which is the case in our example. To do so, we use projection search between the two graphs.

Definition 5.8. *Let $u = (C_u, R_u, E_u, l_u)$ and $v = (C_v, R_v, E_v, l_v)$ be two basic conceptual graphs defined on the same vocabulary V. A projection of v in u is a function $P : V \times V \rightarrow (C_u \times C_v) \cup (R_u \times R_v)$ of the nodes such that the arcs and their labels are preserved and the labels of the nodes can be specialized.*

- $\forall (r_u, i, c_u) \in u, (P(r_u), i, P(c_u)) = (r_v, i, c_v) \in v$
- $\forall e \in C_u \cup R_u, l_v(P(e)) \leq l_u(e)$

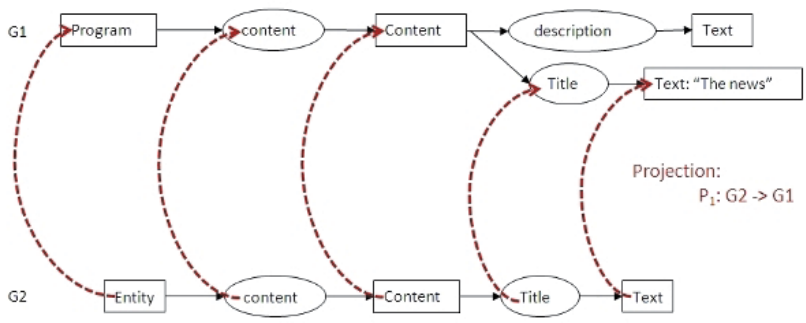


Fig. 12. Example projection between two graphs

Figures 12 depicts an example of projection. $G2$ can be projected in the graph $G1$ through the projections P_1 and $G1$ is more specific than $G2$. We use injective projections. Two different nodes of one graph have two different images in the other graph. Maximal join is a projection based operation defined on conceptual graphs.

The maximal join is composed of two components. First, it tests the compatibility of two elements of the graphs and then fuses them actually. To define the maximal join operation, Sowa defines several other operations that we detail hereafter.

Definition 5.9. *If a conceptual graph u is canonically derivable (see Sowa (2000)) from a conceptual graph v, then u is called a specialization of v and v is called a generalization of u.*

Definition 5.10. *Let two conceptual graphs $u1$ and $u2$ have a common generalization v with injective projections $P1 : v \rightarrow u1$ and $P2 : v \rightarrow u2$. $P1$ and $P2$ are* compatible projections *if, for each concept c in v, the following conditions are true:*

- *$P1(c)$ and $P2(c)$ have a common subtype,*
- *the referents of $P1(c)$ and $P2(c)$ are either equal or one of them is undefined.*

The definition of the maximal join of two graphs u1 and u2 given by Sowa in Sowa (1984) is the following one.

Definition 5.11. *Let v be the most specific common generalization of the graphs $u1$ and $u2$. There is no generalization $v2$ of $u1$ and $u2$ such as v is a sub-graph of $v2$.*
$P1$ and $P2$ are two compatible injective projections of v in $u1$ and $u2$. $P1$ and $P2$ are maximally extended ($P1$ and $P2$ are maximally extended if they have no extension).
A join on these projections is called a maximal join.

There may be several possibilities of fusion between two observations, according to which combinations of observed items are fused or not. This phenomenon is well manage by the maximal join operator. As there may exist several maximally extended compatible projections between two graphs, joining two graphs maximally may give several results, each one of them being a fusion hypothesis.
However, using the maximal join only is not sufficient in order to fuse information as it enables to fuse only strictly identical values. Figure 13 gives an example of such a case. Domain knowledge must used in order to extend the notion of compatibility between concepts, so that concepts with sufficiently similar referents can be be fused.

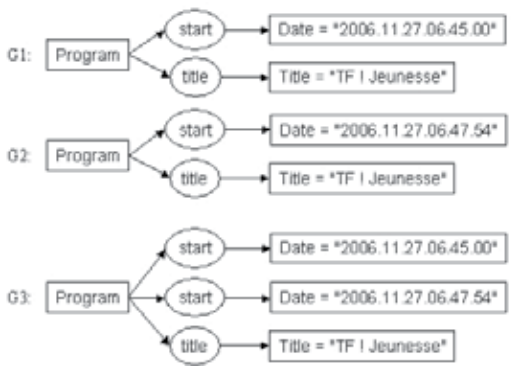


Fig. 13. Limitations of Maximal Join (1)

To do so, we use *Fusion Strategies* which are rules encoding domain knowledge and fusion heuristics.

5.4.2 Fusion strategies

Formally, the fusion strategies are expressed as rules that encompass two functions: a *compatibility testing function* that we call $\mathrm{f_{comp}}$, and a *fusion function* that we call $\mathrm{f_{fusion}}$.
For one application, we define a set of fusion strategies, each one associated with a (set of) conceptual type(s) for which the strategy is valid.

Definition 5.12. *Let c_1 and c_2 be concepts defined on the same vocabulary such that $c_1 \in G_1$ and $c_2 \in G_2$. A fusion strategy* $\mathrm{strategy_{fusion}}$ *is defined as follows:*

$$\begin{aligned} strategy_{fusion} = \ & if \ \ \mathrm{f_{comp}}(c_1, c_2) \\ & then \ \mathrm{f_{fusion}}(c_1, c_2) \\ & else \ \{c_1, c_2\} \end{aligned}$$

where $\mathrm{f_{comp}} : \mathcal{C} \times \mathcal{C} \rightarrow \{true, false\}$ *is a function testing the compatibility of two concept nodes, and* $\mathrm{f_{fusion}} : \mathcal{C} \times \mathcal{C} \rightarrow \mathcal{C}$ *is a fusion function upon the concepts nodes of the graphs.*

The fusion strategies applied on two concept nodes result either in a fused concept node if the initial nodes are compatible, or in the initial nodes themselves if they are incompatible.

5.4.3 Definition of the compatibility function

A compatibility function $\mathrm{f_{comp}} : \mathcal{C}^2 \rightarrow \{true, false\}$ is defined regarding the similarity that exists between two concepts.

Definition 5.13. *Let $(c_1, c_2) \in \mathcal{C}^2$, with t_1 (respectively t_2), the conceptual type of c_1 (respectively c_2) and v_1 (respectively v_2) the referent of c_1 (respectively c_2).*
The compatibility function $\mathrm{f_{comp}}$ *is then defined as follows:*

$$\begin{aligned} \mathrm{f_{comp}}(c_1, c_2) &= \mathrm{sim_{fusion}}\, c_1, c_2 \geq t \\ &= \mathrm{sim_{type}}(t_1, t_2) * \mathrm{sim_{ref}}(v_1, v_2) \geq t \end{aligned}$$

Where $t \in [0, 1]$ is a threshold defined by the domain expert, $\mathrm{sim_{ref}}$ *is one of the similarity measures defined for referents earlier and* $\mathrm{sim_{type}}(t_1, t_2)$*defined as follows.*

To decide whether two concepts can be fused, regarding their conceptual types, one has to check whether the conceptual types have a common subtype. The fusion precises the observations. Regarding the fusion of conceptual types, this means that when fused, two conceptual types will result in their most general common subtype if this sub type exist.

Definition 5.14. *The similarity* $\mathrm{sim_{type}} : T_C^2 \rightarrow [0, 1]$ *is defined as follows:*
$\forall (t_1, t_2) \in T_C \times T_C$,

$$\mathrm{sim_{type}}(t_1, t_2) = \begin{cases} 1 \ if \ \exists t \in \mathcal{V} \ such \ that \ t \leq t_1 \ and \ t \leq t_2 \\ 0 \ otherwise. \end{cases}$$

As a matter of example, considering the fusion of TV program descriptions, let $c_1 = [t_1 : v_1]$ and $c_2 = [t_2 : v_2]$ where the most general common subtype of t_1 and t_2 is $t_{1,2}$.
The compatibility between two `Text` concepts is tested thanks to the following compatibility function:
$\mathrm{f_{comp}}_{\{\text{"}Text\text{"}\}}(c_1, c_2) = \mathrm{sim_{type}}(t_1, t_2) * \mathrm{sim_{ref}}_{strings}(v_1, v_2) > 0,8$
With $\mathrm{sim_{type}}$ and $\mathrm{sim_{ref}}_{strings}$ defined above.

The compatibility between two `Date` concepts is given by the following function:
$fcomp_{\{\text{"}Date\text{"}\}}(c_1, c_2) =$
$\text{sim}_{\text{type}}(t_1, t_2) * \text{sim}_{\text{ref}num}(v_1, v_2) > 0$
and

- $\text{sim}_{\text{ref}num}(v_1, v_2) = 0$ if $|v_1 - v_2| \geq 5$
- $\text{sim}_{\text{ref}num}(v_1, v_2) = 1 - \frac{|d_1 - d_2|}{5}$ otherwise

With sim_{type} defined in 5.14, page 20 and $\text{sim}_{\text{ref}num}$ defined above.
v_1 and v_2 are numeric values expressing the dates in numbers of minutes. For instance, the 27th of November 2006 at 6.45 am, written "2006.11.27.06.45.00" in the figures depicting the example graphs is expressed as: 200611270645.

5.4.4 Definition of the fusion function

The fusion function allows, for any couple of concept nodes, to process, if it exists, the concept node resulting from the fusion of the two initial nodes:

Definition 5.15. *The function* $\text{f}_{\text{fusion}} : \mathcal{C}^2 \rightarrow \mathcal{C}$ *is defined as follows:*

$$\text{f}_{\text{fusion}}(c_1, c_2) = c$$

where $c \in \mathcal{C}$ *is the concept that results from the fusion of* c_1 *and* c_2.

For instance, when fusing two "Text" concepts, we may choose to keep the longest of the two compatible string values.

5.4.5 Extension of the maximal join operation

The fusion strategies are used to extend the maximal join operation that was initially defined by Sowa. The notion of compatibility between two concept nodes is extended and the construction of the joint (i.e. fused) concepts is also modified, allowing to use the fusion function. We call this extension "*maximal join given a fusion strategy*".

Fig. 14. Compatible concepts given a fusion strategy

Two concepts are compatible (Figure 14) if

- they have a most general common sub-type, and
- their values conform this most general common sub-type,
- they are compatible given the selected compatibility function.

The fusion of two concepts is a concept with their most general common sub-type as type, and the result of the fusion function applied to their values as value.
To process the compatibility of two relation nodes, we consider their types and the neighboring concepts. Their types must be identical, and the concepts must be compatible pair wise, respecting the labels of the edges that link them to the relations.

To compute the extended maximal join of two graphs, we have to find compatible sub-graphs of the two graphs that are maximally extended in terms of the number of their nodes. The compatibility of the two subgraphs is processed according to the compatibility of their concepts and relation nodes. Then, the compatible subgraphs are fused.

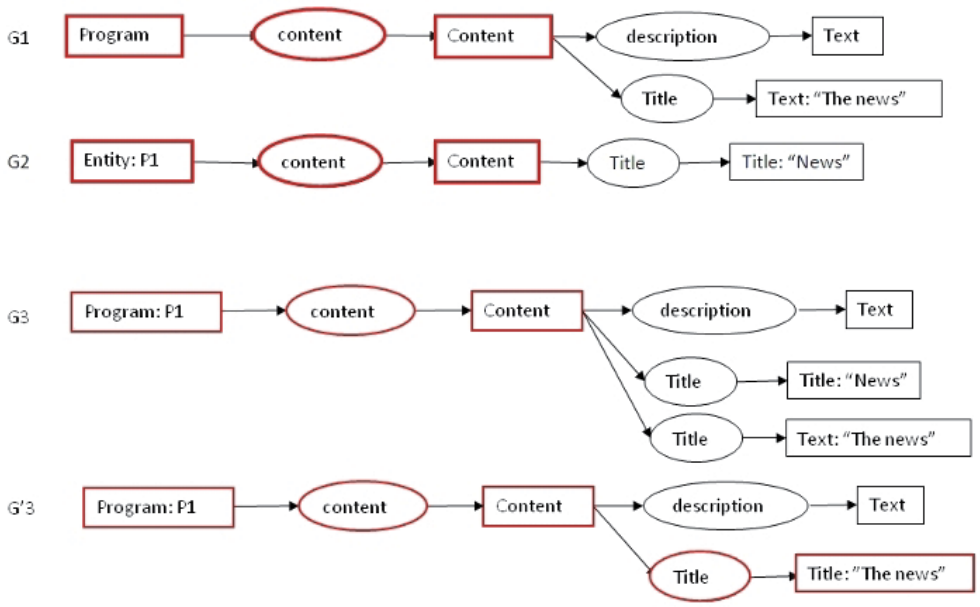


Fig. 15. Compatible relations (Extended maximal join)

On the example depicted on figure 15, we try to fuse the graph $G1$ and $G2$. We can can see that, according to the initial definition of compatibility between concepts and relation nodes, the subgraphs of $G1$ and $G2$ composed of `Program` and `Content` concepts, linked through a `content` relation on the one hand and `Entity` and `Content` concepts, linked through a `content` relation on the other hand are compatible. The result of the maximal join is thus the one depicted on graph $G3$.
When looking at the "News" and "The news" titles of the two programs depicted on figure 15 and given the remaining elements of the two descriptions, one would like to fuse the titles. Indeed, they "obviously" represent the same title and the descriptions are related to the same TV program. By including domain knowledge thanks to the compatibility testing function, we obtain as compatible subgraphs the two previous ones, with the titles in addition. The result of the fusion of $G1$ and $G2$ using maximal join extended with the fusion strategies defined in the examples above gives the graph $G'3$.

6. Experimentations

We describe here experiments that were conducted in order to validate the usefulness of both the fusion strategies and the association phase within soft data fusion. We used information acquired within the TV program case study described earlier. The fusion platform that was developed was also used within biological experimentation and intelligence applications.

6.1 Experimentation protocol

We fused TV program descriptions acquired on the DVB stream of meta data and an on-line TV magazine, from sixteen TV channels during 24 hours.
We used the TV program descriptions provided by the *INAthèque* as reference data to evaluate our fusion. The INA records the descriptions of all the programs broadcast on the french TV and radio. Thereby, we know whether a fused program corresponds to the program that was really played.
For our experimentation, we request every 5 minutes the two sources of information to give us the next program scheduled on one channel. The two provided TV program descriptions are then fused using one of the fusion strategies.
After fusion, we compare the fused TV program descriptions to the INA reference data. If the titles, subtitles, channels etc. are compatible, the fused program description is considered to be correctly found with regards to reality. The results that we obtained are detailed in the following sections.

6.2 Fusion strategies

The quality of the fusion that we obtained using different strategies was measured. To this aim, we launched our experimentations using the fusion platform first combined with no strategy and then with three different ones. The first experiment -no fusion strategy- is equivalent to using the initial maximal join operator for information fusion.
The strategies that encode domain knowledge are the following ones:

Strategy 1 extends dates compatibility. Two dates are compatible if the difference between the two is less than five minutes. If two dates are compatible but different, the fused date should be the earliest one if it is a "begin date" and the latest one otherwise.

Strategy 2 extends dates and titles compatibility. The dates compatibility is the same as for strategy 1. Two titles are compatible if one of them is contained in the other one. If two titles are compatible but different, the fused title should be the longest one.

Strategy 3 extends dates and titles compatibility. The dates compatibility is the same as for strategy 1. Two titles are compatible if the length of the common substrings exceeds a threshold. If two titles are compatible but different, the fused title should be the longest one.

6.3 On the usefulness of fusion strategies

As first interpretation, we compared the percentage of programs that were correctly found after fusion, to the reference data, and looked at the variations resulting of the use of the different strategies. Figure 16 shows the results that we obtained on a representative selection of TV channels. As expected, we can see that the fusion of observations using the maximal join operation only is not sufficient. Only the descriptions with strictly identical values are fused. Applying the three previously cited fusion strategies, we can see that the more the compatibility constraints between two values are relaxed, the better the results are. It is equivalent to inject more and more domain knowledge in the fusion process.
The different experimentations that we carried out showed that the quality of the fusion process is heterogeneous, according to several parameters. One of these parameters on which the fusion results can be dependent, is the period of the day and the specificity of the channel. For non-popular channels (BFM...) and at periods of low audience (early morning), we observed a lot of errors in the programs given by the TV magazine.

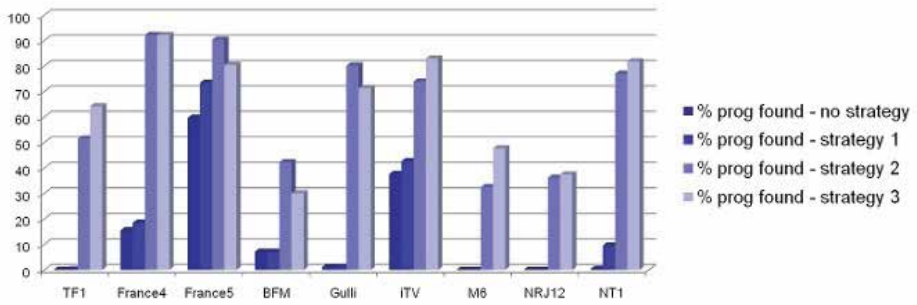


Fig. 16. Percentage of programs correctly fused and identified with different strategies

6.4 On the usefulness of association

The results of the fusion of TV programs that are scheduled on periods of low audience are very bad. Among other particularities, we observed that the TV magazine has "holes", especially for non-popular channels. During such periods, as next program to be broadcast, the magazine source of information gives a program that will actually be broadcast several hours after, whereas, the DVB gives the real next one.

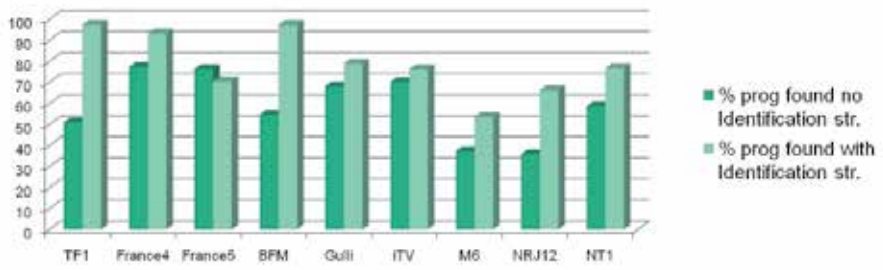


Fig. 17. compatibility testing

The two descriptions are then incompatible and the resulting fused program is not well formed (it has two different titles or begin dates for instance). To overcome such problems, we introduced the use of the association phase.
Figure 17 shows the different percentages of program correctly found, first without association, then using one based on title similarity and distance between begin and end times.

7. Conclusion and future work

We studied the issue of soft data fusion. The aim is to propose a generic approach and a generic framework for high level information fusion. The information items represent the descriptions of (part of) complex situations that themselves contain several actors or objects linked together through semantic relationships.
Besides, the proposed framework enables fusing informations items coming from several sources. Discrepancies among pieces of information are studied and we detect that two lightly different pieces of information concern the description of the same situation, and then choose what the fused description should look like.
We focused on three aspects regarding information fusion: the modeling of the situation of interest, the association phase, which aims at deciding whether two observations concern

the same real world situation or not, and the information synthesis phase, where compatible observations of a single real situation are fused.
Regarding situation modeling, we showed that the conceptual graphs formalism could be used in order to represent situations of interest that have to be monitored.
The association phase relies on the use of similarity measures between graphs structures. Some parts of the measures are generic whatever the application domain is. Other components must be customized either by using specific similarity measures, or thanks to thresholds and weights. The measures we propose take into account the similarity of the values or referents of the concepts.
The information synthesis phase relies on the use of the maximal join operation defined on the conceptual graphs structures. We adapted this operation, that was initially proposed by John Sowa in Sowa (1984), by relaxing the constraints during the similarity testing of two concept nodes in the fusion process. Through this information synthesis step, we tackle the problem of soft data fusion and take into account the issue of discrepancies between the different information items. We use both a compatibility testing and a fusion functions inside the maximal join operation.
Finally, we show the usefulness of our proposition within a real application. We described how we propose to take the advantage of information fusion within a TV program recommendation system.

8. References

Bosse, E., Roy, J. and Wark, S. (2007). *Concepts, Models, and Tools for Information Fusion*, Artech House Publishers.

Buford, J., Lewis, L. and Jakobson, G. (2008). Insider threat detection using situation-aware mas, *In Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, pp. 212–219.

Chein, M. and Mugnier, M.-L. (1992). Conceptual graphs: fundamental notions, *Revue d'Intelligence Artificielle* 6(4): 365–406.

Chein, M. and Mugnier, M.-L. (2008). *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*, Advanced Information and Knowledge Processing, Springer.

de Chalendar, G., Grau, B. and Ferret, O. (2000). Conceptual graphs generalization, *RFIA 2000. 12ème Congrès Francophone AFRIT-AFIA*, Paris.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems, *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37: 32–64(33).
URL: *http://www.ingentaconnect.com/content/hfes/hf/1995/00000037/00000001/art00004*

Gandon, F., Corby, O., Diop, I. and Lo, M. (2008). Distances semantiques dans des applications de gestion d'information utilisant le web semantique, *Proc. Workshop Mesures de similarites semantique, EGC*, INRIA Sophia Antipolis - Mediterranee.

Godbillon-Camus, B. and Godlewski, C. J. (2007). Credit risk management in banks: Hard information, soft information and manipulati, Munich Personal RePEc Archive.
URL: *http://mpra.ub.uni-muenchen.de/*

Hall, D. and Llinas, J. (2001). *Handbook of Multisensor Data Fusion*, CRC Press.

Laskey, K. (2008). Probabilistic ontologies for knowledge fusion, *In Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, pp. 1402–1409.

Laudy, C. (2010). *Introducing semantic knowledge in high level information fusion*, PhD thesis, Université Pierre et Marie Curie.

Matheus, C., Kokar, M. and Baclawski, K. (2003). A core ontology for situation awareness, *6th International Conference on Information Fusion*, Cairns, Queensland, Australia, pp. 545–552.

MIP (2005). Joint c3 information exchange data model (jc3iedm main), *Technical report*, Greding, Germany.

Peirce, C. S. (1932). *Collected papers of Chalres Sanders Peirce*, Vol. 2, C. Hartshone & P. Weiss (Eds.), Cambridge, MA.

Petersen, M. (2004). Information: Hard and soft.
URL: *www.kellogg.northwestern.edu/faculty/petersen/htm/papers/softhard.pdf*

Pravia, M., Prasanth, R., Arambel, P., Sidner, C. and Chong, C.-Y. (2008). Generation of a fundamental data set for hard/soft information fusion, *In Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, pp. 1621–1628.

Sambhoos, K., Llinas, J. and Little, E. (2008). Graphical methods for real-time fusion and estimation with soft message data, *In Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, pp. 1621–1628.

Sorlin, S., Champin, P.-A. and Solnon, C. (2003). Mesurer la similarité de graphes étiquetés, *9èmes Journées Nationales sur la résolution pratique de problèmes NP-Complets*, pp. 325–339.

Sowa, J. (1976). Conceptual graphs for a data base interface, *IBM Journal of Research and Development* 4(20): 336-357.

Sowa, J. F. (1984). *Conceptual Structures. Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA.

Sowa, J. F. (2000). Ontology, metadata, and semiotics, *in* Springer-Verlag (ed.), *ICCS'2000: Conceptual Structures: Logical, Linguistic, and Computational Issues*, Vol. 1867/2000 of *LNAI*, Darmstadt, Germany, pp. 55–81.

2

Knowledge Representation and Validation in a Decision Support System: Introducing a Variability Modelling Technique

Abdelrahman Osman Elfaki[1], Saravanan Muthaiyah[2]
Chin Kuan Ho[2] and Somnuk Phon-Amnuaisuk[3]
[1]*Management and Science University*
[2]*Multimedia University Cyberjaya,*
[3]*Tunku Abdul Rahman University*
Malaysia

1. Introduction

Knowledge has become the main value driver for modern organizations and has been described as a critical competitive asset for organizations. An important feature in the development and application of knowledge-based systems is the knowledge representation techniques used. A successful knowledge representation technique provides a means for expressing knowledge as well as facilitating the inference processes in both human and machines [19]. The limitation of symbolic knowledge representation has led to the study of more effective models for knowledge representation [17].

Malhotra [14] defines the challenges of the information-sharing culture of the future knowledge management systems as the integration of decision-making and actions across inter-enterprise boundaries. This means a decision making process will undergo different constraints. Therefore, existence of a method to validate a Decision Support System (DSS) system is highly recommended. In the third generation of knowledge management, the knowledge representation acts as boundary objects around which knowledge processes can be organized [26]. Knowledge is viewed in a constructionist and pragmatic perspective and a good knowledge is something that allows flexible and effective thinking and construction of knowledge-based artifacts [26].

This paper answers the two questions of [26] and [14] in the context of a DSS: 1) how to define and represent knowledge objects and 2) how to validate a DSS.

For any decision, there are many choices that the decision maker can select from [7]. The process of selection takes place at a decision point and the selected decision is a choice. For example, if someone wants to pay for something, and the payment mode is either by cash or by credit card, the payment mode is the decision point; cash and credit card are choices. Now, we can conclude that the choices and decision points represent the knowledge objects in DSS. Choices, decision points and the constraint dependency rules between these two are collectively named as variability. Task variability is defined in [5] as the number of exceptions encountered in the characteristics of the work. The study in [5] tested the importance of variability in the system satisfaction. Although there are many existing approaches for representing knowledge DSS, the design and implementation of a good and useful method that considers variability in DSS is much desired.

The term variability generally refers to the ability to change; to be more precise, this kind of variability does not occur by chance but is brought about on purpose. In other words, variability is a way to represent choices. Pohl et al. [20] suggest the three following questions to define variability.

What does it vary? : Identifying precisely the variable item or property of the real world. The question leads us to the definition of the term variability subject (A variability subject is a variable item of the real world or a variable property of such an item).

Why does it vary? : There are different reasons for an item or property to vary: different stakeholders' needs, different countries laws, technical reasons, etc. Moreover, in the case of interdependent items, the reason for an item to vary can be the variation of another item.

How does it vary? : This question deals with the different shapes a variability subject can take. To identify the different shapes of a variability subject, we define the term variability object (a particular instance of a variability subject).

Example of variability *Subject and Objects for "Car":*

The variability subject "car" identifies a property of real-world items. Examples of variability objects for this variability subject are Toyota, Nissan, and Proton.

The problem of representing variability in a DSS requires a complex representation scheme to capture static and dynamic phenomena of the choices that can be encountered during the decision process. We believe that the key feature of such knowledge representation (for variability in a DSS) is its capability of precise representation of diverse types of choices and associations within them. This involves: i) qualitative or quantitative description of choices and their classification, ii) representation of causal relationships between choices and iii) the possibility of computerizing the representation.

The main aim of variability representing in DSS is to create a decision repository that contains decision points, its related choices and the constraint dependency relations between decision points-choices, choices-choices, or decision points-decision points.

Nowadays, Feature Model (FM) [12] and Orthogonal Variability Model (OVM) [20] are the well-known techniques to represent variability. Although, FM and OVM are successful techniques for modeling variability, some challenges still need to be considered such as logical inconsistency, dead features, propagation and delete-cascade, and explanation and corrective recommendation. Inconsistency detection is defined as a challenging operation to validate variability in [2]. In [27] the source of logical inconsistency is defined from a skill-based or rule-based errors which would include errors made in touch-typing, in copying values from one list to another, or other activities that frequently do not require a high level of cognitive effort. One of the main drawbacks coming from the fusion of several different and partial views is logical inconsistency [9]. Dead feature is defined in [6] as a frequent case of error in feature model- based variability. Instead of dead feature, we called it dead choice. Modeling variability methods must consider constraint dependency rules to assure the correctness of the decision. Propagation and delete cascade operation is proposed to support auto selection of choices in the decision making process. Propagation and delete cascade operation is a very critical operation in the semi-auto environment.

This paper defines a rule-based approach for representing and validating knowledge in DSS. In addition to representing variability to capture knowledge in DSS, intelligent rules are defined to validate the proposed knowledge representation. The proposed method validates two parts. The first part is validating a decision repository in which a logical inconsistency and dead choices are detected. The second part is validating the decision

making process by providing automated constraint dependency checking, explanation and corrective recommendation, and propagation and delete-cascade.
This paper is structured as follows: Literature is surveyed in section two. Knowledge representation of DSS using variability is demonstrated in section three. Knowledge validation is illustrated in section four and implementation is discussed in section five. Section six contains the conclusion and future work.

2. Related work

The aim of knowledge representation is to facilitate effective knowledge management which concerns expressive representation and efficiency of reasoning in human [15]. Related works on this area are summarized as follows:
Haas [10] investigated the feasibility of developing an overarching knowledge representation for Bureau of Labor Statistics information that captured its semantics, including concepts, terminology, actions, sources, and other metadata, in a uniformly applicable way. Haas suggested the (ISO/IEC 11179) standard for metadata, as knowledge representation techniques. Molina [16] reported the advantages of using knowledge modeling software tool to help developers build a DSS. Molina describes the development of DSS system called SAIDA where knowledge is represented as components, which was designed by Knowledge Structure Manager (KSM). KSM is a knowledge-modeling tool that includes and extends the paradigm of task method-domain followed by different knowledge engineering methodologies. KSM provides a library of reusable software components, called primitives of representation that offer the required freedom to the developer to select the most convenient representation for each case (rules, frames, constraints, belief networks, etc.).
Froelich and Wakulicz-Deja [8] investigated problems of representing knowledge for a DSS in the field of medical diagnosis systems. They suggested in [8] a new model of associational cognitive maps (ACM). The ability to represent and reason with the structures of causally dependant concept is the theoretical contribution of the proposed ACM. Antal [1] proposed the bayesian network as a knowledge representation technique to represent multiple-point-of views. The proposed technique in [1] serves as a refection of multiple points of view and surpasses bayesian network both by describing dependency constraint rules and an auto-explanation mechanism. Lu et al. [13] developed a knowledge-based multi-objective DSS. The proposal in [13] considers both declarative and procedural knowledge. Declarative knowledge is a description of facts with information about real-world objects and their properties. Procedural knowledge encompasses problem-solving strategies, arithmetic and inferential knowledge. Lu et al. [13] used text, tables and diagrams to represent knowledge.
Brewster and O'Hara [3] prove difficulties of representative skills, distributed knowledge, or diagrammatic knowledge using ontologies. Pomerol et.al [21] used artificial intelligence decision tree to represent operational knowledge in DSS. Christiansson [4] proposed semantic web and temporal databases as knowledge representation techniques for new generation of knowledge management systems. One of the most sophisticated knowledge modeling methodologies is Common KADS [24]. Common KADS explains how to model a knowledge application through structural top-down analysis of the problem domain. The outcome of modeling process according to Common KADS consists of three layers that are called contextual model, conceptual model and design model. Common KADS model did not provide mechanism to define relation between objects or between layers. Padma and

Balasubramanie [18] used traditional knowledge-based tool to define DSS. Williams [28] described the benefits of using Ontologies and Argumentation for DSS. Suh in [25] applied Database Management System (DBMS) in two-phased decision support system for resource allocation.

To the best of our knowledge, there is no one particular or specific method for handling variability as a knowledge representation technique in DSS. In addition to variability representation, our proposed method could be used to deal with main challenges in variability representation such as: constraint dependency rules, explanation, propagation and delete-cascade, logic inconsistency detection and dead decision detection. Table 1 summarized the previous works in knowledge representation and validation regarding a DSS. The columns are denoted as following: KR for Knowledge Representation, CDR for Constraint Dependency Rules, Expl for Explanation, Pro and DC for Propagation and delete-cascade, LID for Logic Inconsistency Detection and DDD for Dead Decision Detection.

Technique	Ref.	KR	Reasoning	CDR	Expl	Pro and DC	LID	DDD	Gap
ISO/IEC 11179	10	Yes	No	No	No	No	No	No	6/7
Traditional artificial intelligence knowledge representation techniques such as frames, decision trees, belief networks, etc	16,1,12	Yes	Yes	No	No	No	No	No	5/7
Associational cognitive maps(ACM)	8	Yes	Yes	No	No	No	No	No	5/7
Bayesian network	1	Yes	Yes	Yes	Yes	No	No	No	3/7
Text, tables and diagrams	13	Yes	No	No	No	No	No	No	6/7
Ontologies	3, 28	Yes	Yes	No	No	No	No	No	5/7
Temporal database and Semantic Web	4	Yes	Yes	Yes	Yes	No	No	No	3/7
Three layer modeling(KADS)	24	Yes	Yes	No	No	No	No	No	5/7
DBMS	25	Yes	Yes	No	No	No	No	No	5/7

Table 1. Summary of Literature Review

Table 1 clarifies the need for new method of representation and validating knowledge in DSS. Although the logic inconsistency and dead decision problems are cleared in variability representation methods, some works in literature expressed these problems in a DSS [27].

3. Knowledge representation in DSS using variability

In this section, we defined variability is DSS, and then described how to represent variability in a DSS using First Order Logic (FOL).

3.1 Define variability in DSS

By variability in DSS, we mean choices representation (considering dependency constraint rules between them). In the choice phase, the decision maker chooses a solution to the problem or opportunity. DSS help by reminding the decision maker what methods of choice are appropriate for the problem and help by organizing and presenting the information [7]. Hale [11] states that "relationships between knowledge objects such as kind-of, and part-of become more important than the term itself". We can look for choices as knowledge objects. The proposed method defines and deals with these types of relationship and with dependency constraints between choices such as require and exclude.
There is no standard variability representation for a DSS [21]. The proposed method can enhance both readability and clarity in representation of variability in DSS. Consequently, the proposed method represents variability in high degree of visualization (using graph representation) considering the constraints between choices. As we mentioned before in the definition of variability, there are two items: variability subject and variability object. We suggest variability subject as a *decision point* and variability object and a *choice*. As example from figure 1: the variability subject "Promotion" identifies a *decision point*. Examples of variability objects for this variability subject are "Experience, Qualifications, or Special Report". This example illustrated three choices: "Experience, Qualifications, or Special Report" that the decision maker can select from in the decision point "Promotion".

A reward system as an example:

Rewards systems can range from simple systems to sophisticated ones in which there are many alternatives. It is closely related to performance management. Both rewarding and performance measuring are difficult tasks due to the decision variability that takes place in different activities of human resources cycle as it can be seen in figure 1.

3.2 Representing variability in DSS using first order logic

In this sub-section, the notations of the proposed method are explained. Syntaxes and semantics (the most important factors for knowledge representation methods) for the proposed method are defined. The proposed methods composed of two layers. The upper layer is a graphical diagram. Providing visual picture is the function of the upper layer. The lower layer is a representation of the upper layer in forms of predicates. Providing a reasoning tool is the aim of the lower layer. You can imagine the upper layer as a user-interface while the lower layer as a source code. In the lower layer, decision point, choices, and constraint dependency rules are represented using predicates. The output of this process is a complete modeling of variability in DSS as knowledge- based. In other words, this process creates a decision-repository based on two layers. This decision-repository contains all decisions (choices) grouped by decision points. The proposed method validates both decision-repository and decision making process.

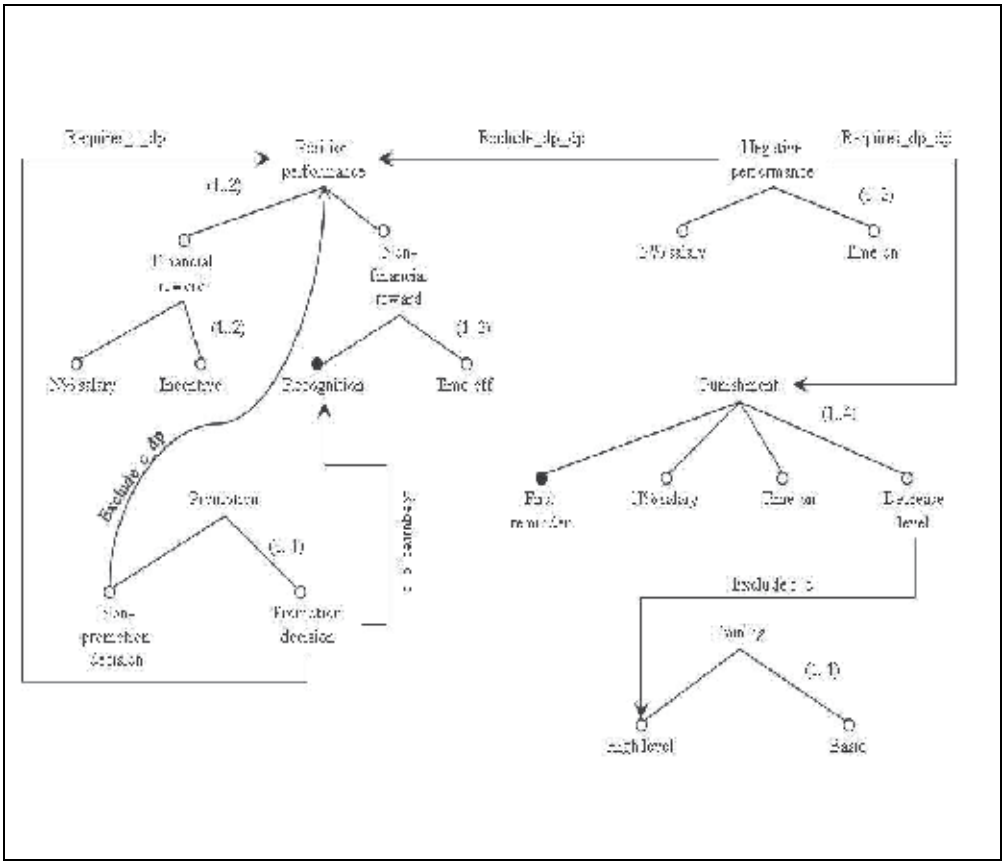


Fig. 1. Variability in Reward System: Upper layer representation

3.1.1 Upper layer representation of the proposed method (FM-OVM)

In this layer, we combined FM diagram with OVM notations. Figure 1 represents the upper layer of our proposed method. Optional and mandatory constraints are defined in Figure 1 by original FM notations [9], and constraint dependency rules are described using OVM notations. OVM and FM can easily become very complex for validating a medium size system, i.e., several thousands of decision points and choices are needed. This reason motivates us to develop an intelligent method for representing and validating variability in DSS.

3.1.2 Lower layer of the proposed method

Decision points, choices, and constraint dependency rules are used to describe variability [20]. Constraint dependency rules are: decision point requires or excludes decision point, choice requires or excludes choice, and choice requires or excludes decision point. In this sub-section, decision points, choices, and dependency constraint rules are described using predicates as a low level of the proposed method. Examples are based on Figure 1. Terms beginning with capital letters represent variables and terms beginning with lower letter represent constants. Table 2 shows the representation of *Negative Performance* decision point using the lower layer of the proposed method.

DECISION POINT

A decision point is a point that the decision maker selects one or more choices. In figure 1, *training* represents a decision point and *Basic* and *High level* represent its choices. The following five predicates are used to describe each decision point:

type:

Syntax: type(Name1,decisionpoint).
Semantic: Identify the type, *Name1* defines the name of decision point, e.g. type (promotion, decisionpoint).

choiceof:

Syntax: choiceof(Name1, Name2).
Semantic: Identifies the choices of a specific decision point. *Name1* represents name of decision point and *Name2* represents the name of a choice, e.g. choiceof (promotion, promotion decision)

max:

Syntax: max(Name, int).
Semantic: Identifies the maximum number of allowed selection at a decision point. *Name* defines the name of the decision point and *int* is an integer,e.g. max(positive performance, 2).

min:

Syntax: min(Name, int).
Semantic: Identifies the minimum number of allowed selection at a decision point. *Name* defines the name of the decision point and *int* is an integer, e.g. min(positive performance,1). The common choices(s) in a decision point is/are not included in maximum-minimum numbers of selection.

Common:

Syntax: common(Name1, yes). common(Name2, no).
Semantic: Describes the commonality of a decision point. *Name1* and *Name2* represent the names of decision points, e.g. common(promotion, yes) If the decision point is not common, the second slot in the predicate will become no e.g. common(punishment, no).

CHOICE

A choice is decision belonging to a specific decision point. For instance, in Figure 1, time *on* is a choice that belongs to the *negative performance* decision point. The following two predicates are used to describe a choice

Type:

Syntax: type(Name1,choice).
Semantic: Define the type. *Name1* represents the name of variant, e.g. type(recognition, choice).

Common

Syntax: common(Name1, yes). common(Name2, no).
Semantic: Describes the commonality of a choice, e.g. common(first reminder, yes). If the choice is not common, the second slot in the predicate will become no -as example- common(time on ,no).

Constraint dependency rules

The following six predicates are used to describe constraint dependency rules:

requires_c_c:

Syntax: requires_c_c(Name1, Name2).
Semantic: choice requires another choice. *Name1* represents the first choice and *Name2* represents the required choice, e.g. requires_c_c(promotion decision, recognition).

excludes_c_c:

Syntax: excludes_c_c (Name1, Name2).
Semantic: choice excludes choice. *Name1* represents the first choice and *Name2* represents the excluded choice, e,g. excludes_c_c(decrease level, high level).

requires_c_dp:

Syntax: requires_c_dp(Name1, Name2).
Semantic: Choice requires decision point. *Name1* represents the choice and *Name2* represents the required decision point, e.g. requires_c_dp(promotion decision, positive performance).

excludes_c_dp:

Syntax: excludes_c_dp(Name1, Name2).
Semantic: Choice excludes decision point. *Name1* represents the choice and *Name2* represents the excluded decision point, e.g. excludes_c_dp(non promotion decision, positive performance).

requires_dp_dp:

Syntax: requires_dp_dp(Name1, Name2).
Semantic: Decision point requires another decision point. *Name1* represents the first decision point and *Name2* represents the required decision point, e.g. requires_dp_dp(negative performance, punishment).

excludes_dp_dp:

Syntax: excludes_dp_dp(Name1, Name2).
Semantic: Decision point excludes another decision point. Name1 represents the first decision point and *Name2* represents the excluded decision point, e.g. excludes_dp_dp(negative performance, positive performance)

type(negative performance, decisionpoint).
choiceof(negative performance, time on).
choiceof(negative performance, N% salary).
common(negative performance, no).
min(negative performance, 1).
max(negative performance, 2).
requires_dp_dp(negative performance, punishment).
excludes_dp_dp(negative performance, positive performance).

Table 2. Representation of Negative Performance

In addition to these predicates, we define two more predicates *select* and *notselect*. The *select* predicate is assigned for all selected choices. The *notselect* predicate prevents the choice from

the selection process, i.e. validate decision-making process. At the end, the choices that represent the final decisions are assigned by *select* and not(*notselected*) predicates.

4. Variability validation in DSS

Although variability is proposed as a technique of knowledge representation that provides a decision repository; validating this repository and decision making process is important.
In a decision making process, a decision maker selects the choice(s) from each decision point. The proposed method guides the decision maker by: 1) validating the constraint dependency rules, 2) automatically selecting (propagation and delete-cascade) decisions, and 3) provide explanation and corrective recommendation. In addition, the proposed method validates the decision-repository by detecting dead choices and logical inconsistency. In this section, six operations are illustrated. These operations are implemented using Prolog [29].

4.1 Validating the decision making process

4.1.1 Constraint dependency satisfaction

To validate the decision-making process, the proposed method triggers rules based on constraint dependencies. Based on the constraint dependency rules, the selected choice is either accepted or rejected. After that, the reason for rejection is given and correction actions are suggested. When the decision maker selects a new choice, another choice(s) is/are assigned by the *select* or *notselect* predicates. As example, in table 3: number 1, the choice *x* is

1. $\forall x, y$: type(x, choice) $\wedge$ type(y, choice) $\wedge$ requires_c_c(x, y) $\wedge$ select(x) $\Longrightarrow$ select(y)
2. $\forall x, y$: type(x, choice) $\wedge$ type(y, choice) $\wedge$ excludes_c_c(x ,y) $\wedge$ select(x) $\Longrightarrow$ notselect(y)
3. $\forall x, y$: type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$ requires_c_dp(x, y) $\wedge$ select(x) $\Longrightarrow$ select(y)
4. $\forall x, y$: type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$ excludes_c_dp(x, y) $\wedge$ select(x) $\Longrightarrow$ notselect(y)
5. $\forall x, y$: type(x, decisionpoint) $\wedge$ type(y, decisionpoint) $\wedge$ requires_dp_dp(x, y) $\wedge$ select(x) $\Longrightarrow$ select(y)
6. $\forall x, y$: type(x,decisionpoint) $\wedge$ type(y, decisionpoint) $\wedge$ excludes_dp_dp(x, y) $\wedge$ select(x) $\Longrightarrow$ notselect(y)
7. $\forall x, y$: type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$ select(x) $\wedge$ choiceof (y, x) $\Longrightarrow$ select(y)
8. $\exists x\ \forall y$:type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$ select(y) $\wedge$ choiceof (y, x) $\Longrightarrow$ select(x)
9. $\forall x, y$: type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$ notselect(y) $\wedge$ choiceof (y, x) $\Longrightarrow$ notselect(x)
10. $\forall x, y$: type(x,choice) $\wedge$ type(y, decisionpoint) $\wedge$ common(x,yes) $\wedge$ choiceof (y, x) $\wedge$ select(y) $\Longrightarrow$ select(x)
11. $\forall y$: type(y, decisionpoint) $\wedge$ common(y) $\Longrightarrow$ select(y)
12. $\forall x, y$: type(x, choice) $\wedge$ type(y, υ decisionpoint) $\wedge$ choiceof (y, x) $\wedge$select(x) $\Longrightarrow$ sum(y,(x)) $\leq$ max(y,z)
13. $\forall x, y$: type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$choiceof(y, x) $\wedge$select(x) $\Longrightarrow$ sum(y,(x)) $\geq$ min(y,z)

Table 3. The main rules in the proposed method

selected (assigned by *select* predicate (select(x)) and *x* requires the choice *y*. Then the system automatically assigns *y* by *select* predicate (select(y)). That means *y* is selected also. At the end, *select* and not *notselect* predicates represent the selections of decision-making process.
Table 3 shows the main rules in our proposed method. The proposed method contains thirteen main rules to validate the decision-making process. Rules 1 through 6 are used to validate constraint dependency rules. Rules 7 and 8 deal with relationships between decision point and their choice(s). Rules 10 and 11 satisfy the common property of decision points and choices. Rules 12 and 13 validate the maximum and minimum property of decision points. Appendix 1 describes the proposed rules in details.

4.1.2 Propagation and delete-cascade

This operation defines how some choices are selected automatically as a reaction to previous selection of other choices.
Definition 1: Selection of the choice *n*, *select(n)*, is propagated from selection of the choice *x*, *select(x)*, in three cases:

i. $\forall x,y,z,n{:}type(x,choice) \wedge choiceof(y,x) \wedge select(x) \wedge requires\_dp\_dp(y,z) \wedge type(n,choice) \wedge choiceof(z,n) \wedge common(n,yes) \Longrightarrow select(n).$

If *x* is a choice and *x* belongs to the decision point *y* and *x* is selected, that means *y* is selected (rule 7), and the decision point *y* requires a decision point *z*, that means *z* is also selected (rule 5), and the choice *n* belongs to the decision point *z* and the choice *n* is common. It means the choice *n* is selected (rule 10).

ii. $\forall$ x,n: type(x,choice) $\wedge$ type(n,choice) $\wedge$ select(x)$\wedge$ requires_c_c(x,n) $\Longrightarrow$ select(n).

If the choice *x* is selected and it requires the choice *n*, it means the choice *n* is selected, (rule 1). The selection of choice *n* propagated from the selection of *x*.

iii. $\forall x,z,n{:}type(x,choice) \wedge select(x) \wedge type(z,decisionpoint) \wedge requires\_c\_dp(x,z) \wedge type(n,choice) \wedge choiceof(z,n) \wedge common(n) \Longrightarrow select(n).$

If the choice *x* is selected and it requires the decision point z, that means the decision point *z* is selected (rule 3), and the choice *n* is common and belongs to the decision point *z* and that means the choice *n* is selected (rule 10). The selection of the choice *n* is propagated from the selection of *x*.

Delete-cascade operation

This operation validates the automated decision-making process during execution time. The following scenario describes the problem:
If choice *x* is selected in time *N* and the two choices *y* and *k* are propagated due to selection of *x*, then the decision list (at time *N*) = {*x*, *y*, *k*}. In time (*N* + 1), the choice *m* is selected, and *m* excludes *x*, then *x* is removed from the decision list. The decision list at time (*N* + 1) = {*m*, *y*, *k*}. The presence of the choices *y* and *k* is not correct. The choices *y* and *k* are not decision maker's choices. The following rule implements the delete-cascade operation.

$\forall x,y{:}type(x,choice) \wedge type(y,choice) \wedge requires\_c\_c(y,x) \wedge select(x) \wedge notselect(y) \Longrightarrow notselect(x).$

For all choices *x*, and *y*; if the choice *y* requires *x* and *x* is selected and *y* is assigned by *notselect* predicate, that means *y* is excluded in the configuration process, and *x* was selected according to selection of *y (y requires x)*, then the presence of *x* after exclusion of *y* is not true. The output for this operation is the assignment of the choice *x* with *notselect* predicate. This assignment permits the proposed method to perform *delete-cascade* operation to verify the selections.

4.1.3 Explanation and corrective recommendation

This operation is defined (in this paper) for highlighting the sources of errors within decision- making process. The errors happened due to dissatisfaction of constraint dependency rules.

The general pattern that represents failure due to dissatisfaction of constraint dependency rules is:

Decision A excludes Decision B and Decision A is selected then Decision B fails to select.

In the proposed method, there are two possibilities for the decision: decision point or choice. Three possibilities for the exclusion constraint: choice excludes choice, choice excludes decision point, or decision point excludes decision point. We assign the predicate *notselect* to the excluded choice for preventing future select.

The following definition describes these possibilities in the form of rules:

Selection of choice *n*, *select (n)*, fails due to selection of choice *x*, *select(x)*, and assign by *notselec*t predicate in three cases:

i. $\forall x,y,n{:}type(x,choice) \wedge select(x) \wedge type(y,decisionpoint) \wedge choiceof(y,x) \wedge type(n,choice) \wedge excludes\_c\_dp(n,y) \Longrightarrow notselect(n).$

If the choice *x* is selected, and it belongs to the decision point *y*, this means *y* is selected (Rule 7), and the choice *n* excludes the decision point *y*, this means *n* is assigned by *notselect* predicate.

ii. $\forall x,y,z,n{:}type(x,choice) \wedge select(x) \wedge type(y,decisionpoint) \wedge type(z,decisionpoint) \wedge type(n,choice) \wedge choiceof(y,x) \wedge choiceof(z,n)\ excludes\_dp\_dp(y,z) \Longrightarrow notselect(n).$

If the choice *x* is selected and *x* belongs to the decision point *y*, that means *y* is selected (Rule 7), and if the decision point *y* excludes the decision point *z*, this means *z* is assigned by *notselect* predicate (rule 6), and the choice *n* belongs to decision point *z*, this means *n* is assigned by *notselect* predicate (rule 9).

iii. $\forall x,n{:}\ type(x,choice) \wedge select(x) \wedge type(n,choice) \wedge excludes\_c\_c(x,n) \Longrightarrow notselect(n).$

If the choice *x* is selected, and *x* excludes the choice *n*, which means *n* is assigned by notselect predicate (rule 2).

Two examples are presented to express how the proposed method could be used for guiding the decision maker by explanation and corrective recommendation. Example 1 shows an interactive corrective recommendation mechanism. Example 2 shows how the proposed method validates decision maker in future based on his current selections.

Example 1

Suppose the decision maker selected *decrease level* as a punishment for one employee. After that, the decision maker selects *high level* as training for the same employee; the system rejects the choice and directs the decision maker to deselect *decrease level* first. Table 4 describes Example 1. This example represents rule (iii). The example illustrates how the proposed method guides decision makers to solve the rejection reason.

Example 2

The decision maker asks to select the choice *non promotion decision*, which is excludes *positive performance* decision point. The system accepts the choice and assigns the decision point *positive performance* by *notselect* predicate to validate future selections. Table 5 describes Example 2. The predicate *notselect (positive performance)* prevents the selection of its choices, Rule 9.

The proposed method guides the decision maker step by step (in each choice). If the decision maker's choice is invalid, the choice is immediately rejected and corrective actions are suggested, see Example 1. Moreover, *notselect* predicate can be assigned to some choices according to decision maker's selection, see Example 2. The *notselect* predicate prevents the decision maker from future errors; see Table 3: Rule 9.

? select (decrease level). You have to deselect high level

Table 4. Example 1

? select (non promotion decision). Yes notselect (positive performance) added to knowledge base.

Table 5. Example 2

4.2 Validate decision repository

4.2.1 Logical inconsistency detection

Inconsistency occurs from contradictions in constraint dependency rules. It is a very complicated problem. Inconsistency has different forms and it can occur between: groups (as example: (A and B) require (D and C) and (D and C) exclude (A and B)), group and individual (as example: (A and B) require D and D excludes (A and B)), or between individuals only (as example**:** (A require**s** B and B requires C and C excludes A)). A, B, C and D could be choices or decision points.

In this paper, we suggest rules to detect logical inconsistency between individuals. The rules that can be used to detect logical inconsistency (between individuals) are categorized in three groups. Each group contains two rules.

Group 1

In this group, we discuss the constraint dependency relation between two decisions from the same type (decision point or choice).

The first decision requires the second one while the second decision excludes the first one. The logical inconsistency between two decisions could be indirect, e.g. A requires B and B requires C and C excludes A. Therefore, to transfer the logical inconsistency to be directly within two decisions, we define these transfer rules:

i. $\forall x,y,c{:}type(x,choice) \wedge type(y,choice) \wedge type(c,choice) \wedge requires\_c\_c(x,y) \wedge requires\_c\_c(y,c) \Longrightarrow requires\_c\_c(x,c)$.

ii. $\forall x,y,c{:}type(x,decisionpoint) \wedge type(y,decisionpoint) \wedge type(c,decisionpoint) \wedge requires\_dp\_dp(x,y) \wedge requires\_dp\_dp(y,c) \Longrightarrow requires\_dp\_dp(x,c)$.

The following rules detect inconsistency in group 1:

i. $\forall x,y{:}type(x,choice) \wedge type(y,choice) \wedge requires\_c\_c(x,y) \wedge excludes\_c\_c(y,x) \Longrightarrow error$.

If the choice x requires the choice y which means selection of x leads to selection of y (Rule 1). In addition, choice y excludes the choice x which means if y selected, x must not be selected (Rule 2). This is an error.

ii. $\forall x,y{:}type(x,decisionpoint) \wedge type(y,decisionpoint) \wedge requires\_dp\_dp(x,y) \wedge excludes\_dp\_dp(y,x) \Longrightarrow error.$

If the decision point x requires the decision point y that means selection of x leads to selection of y (Rule 5), and the decision point y excludes the decision point x which means if y is selected x must not be selected (Rule 6). This is an error.

Group 2

In this group, we discuss the constraint dependency relation between two decision points. At the same time, there is a contradictory relation between one choice (belongs to the first decision point) and the second decision point. The following rules illustrated group 2:

i. $\forall x,y,z{:}type(x,choice) \wedge common(x,yes) \wedge type(y,decisionpoint) \wedge choiceof(y,x) \wedge type(z,decisionpoint) \wedge requires\_dp\_dp(y,z) \wedge excludes\_c\_dp(x,z) \Longrightarrow error.$

If the common choice x belongs to the decision point y, and x excludes the decision point z, which means if x selected, no choice belonging to z must be selected (Rule 4, and Rule 9), and the decision point y requires the decision point z which means if y is selected z must also be selected(Rule 5). Selection of the decision point y means selection of the common choice x (Rule 10) but x excludes z. This is an error.

ii. $\forall x,y,z{:}type(x,choice) \wedge type(y,decisionpoint) \wedge choiceof(y,x) \wedge type(z,decisionpoint) \wedge excludes\_dp\_dp(y,z) \wedge requires\_c\_dp(x,z) \Longrightarrow error.$

If the choice x belongs to the decision point y and x requires the decision point z that means if x selected z should be selected (Rule 3). The decision point y excludes the decision point z that means if one of the choices belongs to y is selected none belongs to z should be selected (Rules 6, 7, and 9). X requires z is an error.

Group 3

In this group, we discuss the constraint dependency relation between two decision points. At the same time, there is a contradictory relation between their choices. The following rules illustrates group 3:

i. $\forall x,y,n,z{:}type(x,choice) \wedge type(y,decisionpoint) \wedge choiceof(y,x) \wedge type(n,choice) \wedge type(z,decisionpoint) \wedge choiceof(z,n) \wedge common(n,yes) \wedge excludes\_c\_c(x,n) \wedge requires\_dp\_dp(y,z) \Longrightarrow error.$

The common choice x belongs to the decision point y and the common choice n belongs to z. The decision point y requires the decision point z that means if y selected then z must be selected, (Rule 5), and selection of y and z means selection of x and n, (Rule 10). X excludes n is an error.

ii. $\forall x,y,n,z{:}type(x,choice) \wedge type(y,decisionpoint) \wedge choiceof(y,x) \wedge type(n,choice) \wedge type(z,decisionpoint) \wedge choiceof(z,n) \wedge requires\_c\_c(x,n) \wedge excludes\_dp\_dp(y,z) \Longrightarrow error.$

If choice x belongs to the decision point y, and the choice n belongs to the decision point z, and x requires n which means if x is selected, n should also be selected (Rule1). Selection of the choice x means selection of the decision point y, and selection of choice n means selection of decision point z (Rule 7). The decision point y excludes the decision point z t which means if one of the choices belonging to y is selected, then none belonging to z must be selected (Rules 6, 7, and 9). X requires n is an error.

4.3 Dead decision detection

A dead decision is a decision that never appears in any legal decision process. The only reason to prevent a decision from being included in any decision process is that there is a

common choice/decision point that excludes this decision. The general form for describing a dead decision is:

Decision A excludes decision B and decision A is common then decision B is a dead decision.

According to the proposed method, there are two possibilities for a decision: decision point or choice, two possibilities for decision (A) to be common and three possibilities for the exclusion.

Two possibilities for decision (A) to be common:

1. Common decision point:
$\exists A$:*type(A, decisionpoint)* $\wedge$ *common(A,yes).*
2. Common choice belongs to common decision point:
$\exists A,C$:*type(A, choice)* $\wedge$ *type(C, decisionpoint)* $\wedge$ *choiceof(C,A)* $\wedge$*common(C,yes)* $\wedge$*common(A,yes)***.**

Three possibilities for the exclusion constraint:

3. Choice excludes choice:
$\exists A,B$: *type(A, choice)* $\wedge$ *type(B, choice)* $\wedge$ *excludes_c_c(A,B).*
4. Choice excludes decision point:
$\exists A,C$:*type(A,choice)* $\wedge$ *type(C, decisionpoint)* $\wedge$ *excludes_c_dp(A,C).*
5. 5. Decision point excludes decision point:
$\exists A,C$: *type(A, decisionpoint)* $\wedge$ *type(C, decisio point)* $\wedge$ *excludes_dp_dp(A,C).*

If we apply the two possibilities of common decision to the three possibilities of the exclusion constraint then we have six possibilities for satisfying the general form of the dead decision. These possibilities are (1, 3),(1, 4),(1, 5),(2, 3),(2, 4),(2, 5). The possibilities (1, 3), (1, 4) and (2, 5) are excluded because decision *A* cannot be decision point and choice at the same time. Therefore, all the possible scenarios for the dead decision are: (1, 5), (2, 3), (2, 4). These scenarios are represented by the following rules:

i. $\forall A,B,C$:*type(A,decisionpoint)*$\wedge$*common(A,yes)*$\wedge$*type(C,decisionpoint)*$\wedge$*excludes_dp_dp(A,C)*$\wedge$ *type(B,choice)* $\wedge$ *choiceof(C ,B)* $\Longrightarrow$ *dead_ decision (B).*

The decision point *C* in the above rule represents a dead decision point. All choices belonging to a dead decision point are dead decisions.

ii. $\forall A,B,C$:*type(A,choice)*$\wedge$*type(C,decisionpoint)*$\wedge$*choiceof(C,A)*$\wedge$*common(C,yes)*$\wedge$*common(A,yes)*$\wedge$ *type(B, choice)* $\wedge$*excludes_c_c(A,B)* $\Longrightarrow$ *dead_ decision(B).*

iii. $\forall A,B,C,D$:*type(A,choice)*$\wedge$*type(C,decisionpoint)*$\wedge$*choiceof(C,A)*$\wedge$*common(C,yes)*$\wedge$*common(A,yes)*$\wedge$*type(B,choice)*$\wedge$*type(D,decisionpoint)*$\wedge$*choiceof(D,B)*$\wedge$*excludes_c_dp(A,D)*$\Longrightarrow$ *dead_ decision (B).*

Rule (i) represents case (1, 5), rule (ii) represents case (2, 3) and rule (iii) represents case (2, 4). The decision point *D* in rule 3 represents a dead decision point. The choice *B* represents all choices belonging to *D*.

5. Scalability testing

Scalability is approved as a key factor in measuring applicability of the techniques dealing with variability modeling [23]. The output time is a measurement key for scalability. A system considers scalable for specific problem if it can solve this problem in a reasonable time.

In this section, we discuss the experiments, which are developed to test the scalability of the proposed method.

5.1 The experiment

In the following, we describe the method of our experiment:

- **Generate the decision repository:** repository is generated in terms of predicates (Decision points, and choices). We generated four sets containing 1000, 5000, 15000, and 20000 choices. Choices are defined as numbers represented in sequential order, as example: In the first set (1000 choices) the choices are: 1, 2, 3,…, 1000. In the last set (20000 choices) the choices are: 1, 2, 3, …, 20000. The number of decision point in each set is equal to number of choices divided by five, which means each decision point has five choices.
- **Define the assumption:** We have three assumptions: i) each decision point and choice has a unique name, ii) each decision point is orthogonal, and iii) all decision points have the same number of choices.
- **Set the parameters**: The main parameters are the number of choices and the number of decision points. The remaining eight parameters (common choice, common decision point, choice requires choice, choice excludes choice, decision point requires decision point, decision point excludes decision points, choice requires decision point, and choice excludes decision point) are defined as a percentage. Three ratios are defined: 10%, 25%, and 50%. The number of the parameters related to choices (such as; common choice, choice requires choice, choice excludes choice, choice requires decision point, and choice excludes decision point) is defined as a percentage of the number of the choices. The number of parameters related to decision point (such as; decision point requires decision point) is defined as a percentage of the number of decision points. Table 6 represents snapshots of an experiment's dataset, i.e. the decision repository in our experiments.
- **Calculate output**: for each set, we made thirty experiments, and calculated the execution time as average. The experiments were done with the range (1000-20000) choices, and percentage range of 10%, 25%, and 50%.

In the following section, the experiments that are done for dead decision detection, explanation, and logical inconsistency detection are discussed. The rest two operations (constraint dependency satisfaction, and propagation and delete-cascade) are working in semi-auto decision environment, where some decisions are propagated automatically according to decisions made. In semi-auto decision environment, the scalability is not a critical issue.

```
type(dp1,decisionpoint).
type(1,choice).
variants(dp1,1).
common(570,yes).
Common(dp123,yes).
requires_c_c(7552,2517).
requires_dp_dp(dp1572,dp1011).
excludes_dp_dp(dp759,dp134).
excludes_c_c(219,2740).
requires_c_dp(3067,dp46).
excludes_c_dp(5654,dp1673).
```

Table 6. Snapshot of experiment's dataset

5.2 Empirical results

Dead Decision Detection: Figure 2 illustrates the average execution time. For (20,000) choices and 50% of constraint dependency rules, the execution time is 3.423 minutes which can be considered as a reasonable time. The output of each experiment is a result file containing the dead decisions.

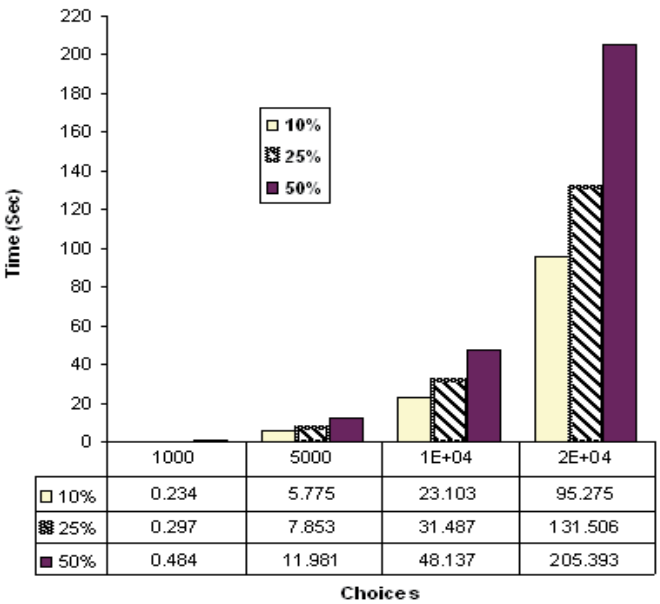


Fig. 2. Dead decision detection results

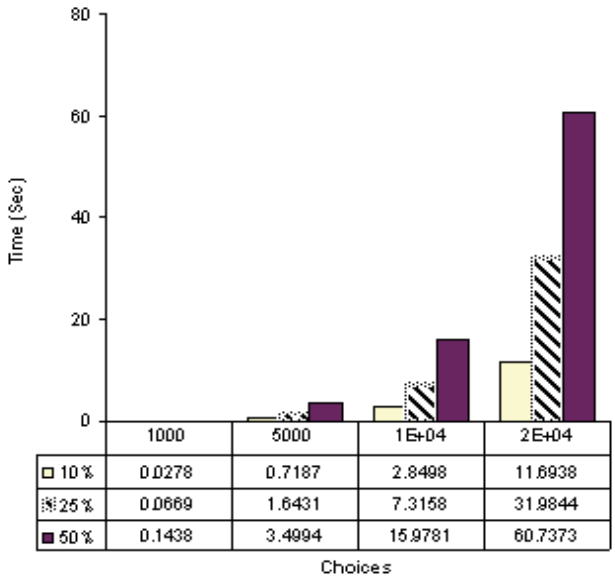
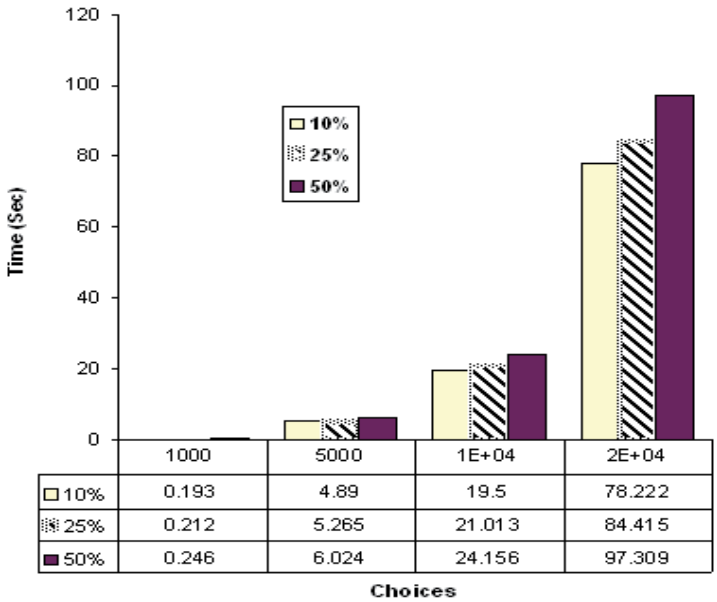


Fig. 3. Explanation results

Explanation

This process defines the source of error that might occur when the new choice is selected. To evaluate the scalability of this operation, we define additional parameter, the predicate select(C): where C is random choice. This predicate simulates decision maker selection. Number of select predicate (defined as a percentage of number of choices) is added to the knowledge-based for each experiment, and the choice C is defined randomly (within scope of choices). Figure 3 illustrates the average execution time. The output of each experiment is a result file containing the selected choices and the directive messages.

Logical Inconsistency-Detection: Figure 4 illustrates the average execution time to detect inconsistency in FM Range from 1000 to 20,000 choices



	1000	5000	1E+04	2E+04
10%	0.193	4.89	19.5	78.222
25%	0.212	5.265	21.013	84.415
50%	0.246	6.024	24.156	97.309

Fig. 4. Logical Inconsistency Detection

6. Conclusion and future work

Representing knowledge objects and the relation between them is the main issues of the modern knowledge representation techniques. We suggest variability for representing knowledge objects in DSS. By introducing variability to represent knowledge in DSS we can get both formalized knowledge representation in decision repository and support decision-making process by validation operations. Decision selection processes are validated using constraint dependency rules, propagation and delete cascade, and explanation and corrective recommendation operations. Decision repository is validated by detecting logical inconsistency and dead choices. In [5] it states, "developing and using a mathematical model in a DSS, a decision maker can overcome many knowledge-based errors". For this reason, the proposed method is supported by FOL rules.

We plan to test and validate this work using real data and real life case studies from industry. In addition, new operations are needed to validate DSS.

7. Appendix A

Explanation of the rules in table 3

Rule 1:

For all choice *x* and choice *y*; if *x* requires *y* and *x* is selected, then *y* is selected.

Rule 2:

For all choice *x* and choice *y*; if *x* excludes *y* and *x* is selected, then *y* is assigned by *notselect* predicate.

Rule 3:

For all choice *x* and decision point *y*; if *x* requires *y* and *x* is selected, then *y* is selected. This rule is applicable as well, if the decision point is selected first and it requires a choice:

$\forall$ x, y: type(x, choice) $\wedge$ type(y, decisionpoint) $\wedge$ require_c_dp(x, y) $\wedge$ select(y) $\Longrightarrow$ select(x)

For all choice *x* and decision point *y*; if *x* requires y and *y* is selected, then *x* is selected.

Rule 4:

For all choice *x* and decision point *y*; if *x* excludes *y* and *x* is selected, then *y* is assigned by *notselect* predicate. This rule is applicable as well, if the decision point is selected first and it requires a choice:

$\forall$ *x, y: type(x, choice)* $\wedge$ *type(y, decisionpoint)* $\wedge$ *exclude_c_dp(x, y)* $\wedge$ *select(y)* $\Longrightarrow$ *notselect(x)*

For all choice *x* and decision point *y*; if *x* excludes *y* and *y* selected, then *x* is assigned by *notselect* predicate.

Rule 5:

For all decision point *x* and decision point *y*, if *x* requires *y* and *x* selected, then *y* is selected.

Rule 6:

For all decision point *x* and decision point *y*, if *x* excludes *y* and *x* is selected, then *y* is assigned by *notselect* predicate.

Rule 7:

For all choice *x* and decision point *y*, where *x* belongs to *y* and *x* is selected, that means *y* is selected.

This rule determines the selection of decision point if one of its choices was selected.

Rule 8:

For all decision point *y* there exists of choice *x*, if *y* selected and *x* belongs to *y*, *x* is selected.

This rule states that if a decision point was selected, then if there is choice(s) belonging to this decision point , must be selected.

Rule 9:

For all choice *x* and decision point *y*; where *x* belongs to *y* and *y* defined by predicate *notselect(y)*, then *x* is assigned by *notselect* predicate. This rule states that if a decision point was excluded, then none of its choices is selected.

Rule 10:

For all choice *x* and decision point *y*; where *x* is a common, *x* belongs to y, and *y* is selected, then *x* is selected. This rule states that if a choice is common and its decision point selected then it is selected.

Rule 11:

For all decision point *y*; if *y* is common, then *y* is selected. This rule states that if a decision point is common then it is selected in any decision process.

Rule 12:

For all choice *x* and decision point *y*; where *x* belongs to *y* and *x* is selected, then the summation of *x* must not be less than the maximum number allowed to be selected from y.

Rule 13:
For all choice x and decision point y; where x belongs to y and x is selected, then the summation of x must not be greater than the minimum number allowed to be selected from y.

Rules 12 and 13 validate the number of choices' selection considering the maximum and minimum conditions in decision point definition (cardinality definition). The predicate sum(y, (x)) returns the summation number of selected choices belongs to decision point y.

8. References

[1] Antal, P." Integrative Analysis Of Data, Literature, And Expert Knowledge By Bayesian Networks",, Phd Thesis , Katholieke University Leuven, 2008.

[2]Batory, D., Benavides, D., Ruiz-Cortés, A., "Automated Analyses of Feature Models: Challenges Ahead", *Special Issue on Software Product Lines , Communications of the ACM*, 49(12) December 2006,pp. 45 – 47.

[3]Brewster, C., O'Hara, K., "Knowledge Representation with Ontologies: The Present and Future", Intelligent Systems, IEEE 1094-7167/04, 19(1), 2004.

[4]Christiansson, P. "Next Generation Knowledge Management Systems For The Construction Industry", Paper w78-2003-80, Proceedings 'Construction It Bridging The Distance. Cib Publication 284. Auckland 2003, pp. 80-87.

[5] Celderman, M. "Task Difficulty, Task Variability and Satisfaction with Management Support systems:Consequences and Solutions", Research Memorandum 1997-53, Vrije University, Amsterdam, 1997.

[6] Czarnecki, K., Hwan, C., Kim, P. "Cardinality-based Feature Modeling and Constraints: A Progress Report", International Workshop on Software Factories at (OOPSLA'05), San Diego California, 2005.

[7] Densham, P. J. "Spatial decision support systems", Geographical information systems: principles and applications, London, Longman, 1991,pp. 403 - 412.

[8] Froelich, W., Wakulicz-Deja, A. " Associational Cognitive Maps for Medical Diagnosis Support", Intelligent Information Systems Conference , Zakopane, Poland, 2008, pp. 387–396.

[9] Grégoire, E. "Logical Traps in High-Level Knowledge and Information Fusion", Specialist Meeting Conference on Information Fusion for Command Support (IST-055/RSM-001), The Hague, The Netherlands,2005.

[10] Haas, S. W. "Knowledge Representation, Concepts, and Terminology: Toward a Metadata Registry for the Bureau of Labor Statistics", Final Report to the United States Bureau of Labor Statistics, School of Information and Library Science, University of North Carolina at Chapel Hill, July 1999.

[11] Hale, P., Scanlan, J., Bru, C. " Design and Prototyping of Knowledge Management Software for Aerospace Manufacturing", 10th ISPE International Conference on Concurrent Engineering, Madeira Island , Portugal, 2003.

[12] Kang, K., Hess, J., Novak, W. Peterson, S. "Feature oriented domain analysis (FODA) feasibility study",Technical Report No. CMU/SEI-90-TR-21, Software Engineering Institute, Carnegie Mellon University, 1990.

[13] Lu, J., Quaddus, M.A., Williams, R. "Developing a Knowledge-Based Multi-Objective Decision Support System", the 33rd Hawaii International Conference on System Sciences, Maui, Hawaii, 2000.

[14] Malhotra, Y. Why knowledge management fails? Enablers and constraints of knowledge management in human enterprises, handbook of knowledge management, chapter 30, springer, 2002, pp. 568-576.

[15] Mikulecky, P., Olsevicov´a, K., Ponce, D. "Knowledge-based approaches for river basin management", *the journal Hydrology and Earth System Sciences*, Discuss., 4, 1999-2033, 2007.

[16] Molina, M. "Building a decision support system with a knowledge modeling tool", *Journal of Decision Systems*, Lavoisier, Volume 14 (3), 2005.

[17] Martinez, S.I. A formalized Multiagent decision support in cooperative environments, Doctoral Thesis, Girona Catalonia, Spain, 2008.

[18] Padma, T., Balasubramanie, P., Knowledge based decision support system to assist work-related risk analysis in musculoskeletal disorder, *Knowledge-Based Systems,* Elsevier, 22, 2009 ,72–78.

[19] Peleg M, Tu S. W.: Decision Support, Knowledge Representation and Management in Medicine, 2006 IMIA Yearbook of Medical Informatics, Reinhold Haux, Casimir Kulikowski, Schattauer, Stuttgart, Germany, BMIR-2006-1088, 2006.

[20] Pohl, k., G. Böckle, Linden, F. J. van der. *Software product line engineering, Foundations, Principles, and Techniques*, Springer, Berlin, Heidelberg, 2005, 490.

[21] Pomerol, J., Brezillon, P. Pasquier, L., "Operational Knowledge Representation for Practical Decision Making", The 34th Hawaii International Conference on System Sciences, 2001.

[22] Roth, B.M., Mullen, J. D. *Decision Making: Its Logic, And practice*, Lanham, MD:Rowman & littlefield, 2002.

[23] Segura S. ,"Automated Analysis of Feature Models using Atomic Sets", the 12th international conference of software product line, Limerick Irland,2008.

[24] Schreiber, G., Akkermans, H., Anjewierden, A., Dehoog, R. Shadbolt, N . Vandevelde, W. Wieling, B, *Knowledge Engineering and Management:The CommonKADS Methodology*, MIT Press, 1999.

[25] Suh C.K. "An Integrated Two-Phased Decision Support System for Resource Allocation", *Wseas transactions on business and economics*,Volume 4(11), November 2007.

[26] Tuomi I., " The Future of Knowledge Management", Lifelong Learning in Europe (LLinE), vol VII, issue 2/2002, , 2002 ,pp. 69-79.

[27] Williams, M., Dennis, A. R., Stam, A., Aronson, J. "The Impact of DSS Use and Information Load on Errors and Decision Quality", Working Papers on Information Systems, Indiana University, USA, Sprouts: Working Papers on Information Systems, 4(22). http://sprouts.aisnet.org/4-22, 2004.

[28] Williams, M. H. Integrating Ontologies and Argumentation for decision-making in breast cancer, PhD Thesis, University College London, 2008.

[29] Wielemaker J.: SWI-Prolog (Version 5.6.36) free software, Amsterdam, University of Amsterdam, 2007.

3

Decision Support Tools for Ontological Engineering

Simon Suigen Guo, Christine W. Chan and Robert Harrison
Energy Informatics Laboratory,
Faculty of Engineering and Applied Science
University of Regina
Canada

1. Introduction

Development cost of knowledge-based systems is often high because knowledge bases are often constructed from scratch in a distributed and heterogeneous environment involving different locations, system types, knowledge representation mechanisms and stakeholders. Hence, it is difficult to share and re-use existing knowledge base components.

Sharing and re-using knowledge can help reduce costs and make it possible to create systems capable of more powerful problem solving. The Semantic Web (SW) is the next evolutionary step for the World Wide Web (Web). The SW aims to provide semantics to data on the Web, enabling computers to more easily share and perform problem solving on the data (Berners-Lee et al. 2001). SW technology can be used to share and re-use knowledge between KBSs in a distributed and heterogeneous environment.

Semantic data networks on the Web provide the bases of the SW, enabling knowledge that is distributed over the Web to be easily shared and processed by a machine. A semantic data structure is required for representing data or knowledge in a shareable format, and an ontology is designed to fulfill this objective. According to the World Wide Web Consortium, an ontology is "a specification of a conceptualization" (Gruber, 1993); ontologies are the vocabularies of the SW that define the concepts and relationships used to describe an area of interest. An ontology represents the definitions of concepts, axioms and facts which discribe real world phenomenon of interest. As suggested in (Gruber, 1993), the declarative formalisms and the set of objects represented as domain knowledge in a knowledge-based system (KBS) can be described by an ontology. Hence, an ontology can also become the basis for building knowledge-based systems. Ontologies implemented in XML[1] based languages, such as RDF[2] and OWL[3], enable different KBS development groups or different Semantic Web applications to share and re-use their knowledge and data.

Knowledge Acquisition (KA) is an important step for building knowledge-based systems (KBS). In the process, software tools are often used for helping knowledge engineers construct ontologies and knowledge bases more efficiently and effectively. The objective of

---

[1] Extensible Markup Language, http://www.w3.org/XML/
[2] Resource Description Framework (RDF), http://www.w3.org/RDF/
[3] OWL Web Ontology Language, http://www.w3.org/TR/owl-features/

our research is to develop a suite of tools that support knowledge visualization and management for the Semantic Web. This suite of tools is created based on the theoretical framework of the Inferential Modeling Technique, which provides support for both static and dynamic knowledge modeling. The tools can be used to support the process of KA during the building of a KBS.

Existing software tools for knowledge engineering support will be critically examined and observations drawn in the assessment will provide the basis for design of the tools we propose for supporting knowledge visualization and management.

This chapter is organized as follows. Section 2 presents background literature relevant to our research. Section 3 describes the development of tools for static and dynamic knowledge visualization and ontology management. Section 4 provides some discussion of the tools. Section 5 provides some conclusions and discussions on directions for future work.

2. Background literature

2.1 Ontology tools

Software tools support knowledge engineers in the difficult task of knowledge acquisition and ontology construction. From a survey of research work on ontology tools, we found the existing tools do not provide adequate support in five areas, which include (Harrison & Chan, 2005): (1) support for an ontological engineering methodology, (2) support for dynamic knowledge modeling, (3) support for dynamic knowledge testing, (4) support for ontology visualization, and (5) support for ontology management. They are discussed in detail as follows:

2.1.1 Lack of support for an ontological engineering methodology

Most of the examined tools were not developed based on an ontological engineering methodology, which can help expedite the ontological engineering process. The methodology provides a sequence of steps, procedure or guidelines that help developers in the process of constructing an ontology. Some examples of research works on tool support for ontological engineering methodologies include OntoEdit which supports On-To-Knowledge (Fensel et al., 2002), WebODE which supports METHONTOLOGY (WebODE), and Knowledge Modeling System (KMS) which supports the Inferential Modeling Technique (IMT) (Chan, 2004).

2.1.2 Lack of support for modeling dynamic knowledge

While a primary function of a tool that supports ontology construction is its ability to model the problem domain. Most tools only model static domain knowledge about classes and attributes but ignore dynamic knowledge, which refers to the knowledge about objectives, tasks and activities (Chan, 2004). A brief survey of the different ontological engineering support tools reveal there are diverse methods employed. For example, the method used by Protégé (Protégé), Ontolingua (Ontolingua), and OntoBuilder (OntoBuilder) for modeling ontologies is to use a hierarchical tree listing all the concepts, and input fields are provided to capture characteristics of concepts. A graphical method of modeling ontologies is employed in tools such as KAON (KANO) and Protégé OWLViz Plug-in (Protégé OWLViz), in which the representational method involves nodes to represent concepts and edges to represent relationships between concepts.

A different approach is adopted in the Unified Modeling Language (UML) tools, which are commonly used to visually describe and communicate the design of software. Due to the similarities between ontologies and object-oriented design, UML class diagrams can be used to model ontology classes and their properties, and relationships between classes (Cranefield et al., 2005). However, UML's expressiveness for modeling ontologies is limited, and Standard UML cannot express more advanced ontologies that contain descriptive logic (Djuric et al., 2005). Due to the limitations of UML, there is on-going work to develop a graphical notation for ontologies, called the Ontology Definition Metamodel (ODM) (Djuric et al., 2005).
The existing tools described can model static knowledge to varying degrees, but they often lack the facilities for modeling dynamic knowledge. For example, it is awkward to link a task to an objective in Protégé using the existing input fields. This is a weakness that our research addresses.

2.1.3 Lack of support for ontology testing

The examined tools that support ontology construction do not include facilities for supporting ontology testing. Analogous to software testing which identifies defects during development and results in a more stable software application being constructed, ontology testing can help the knowledge engineer develop a more complete model of the domain. In the ontological engineering field, ontology testing is also called ontology evaluation. According to (Gomez-Perez et al., 2005), ontology evaluation should be performed on the following items:

- Every individual definition and axiom
- Collections of definitions and axioms stated explicitly in the ontology
- Definitions imported from other ontologies
- Definitions that can be inferred from other definitions

Existing ontology testing systems, such as the OWL Unit Test Framework (OWL Unit Test Framework) and Chimaera's (McGuiness et al., 2000) test suite, evaluate the correctness and completeness of ontologies. Chimaera performs tests in the following four areas: (1) Missing argument names, documentation, constraints, etc., (2) syntactic analysis, e.g. occurrence of words, possible acronym expansion, (3) taxonomic analysis, e.g. unnecessary super classes and types, and (4) semantic evaluation, e.g. slot/type mismatch, class definition cycle, domain/range mismatch (McGuiness et al., 2000). Such testing tools are sufficient for testing static knowledge, but are not suitable for testing the interactions between behaviour and objects.
Our research work aims to fill this gap in research by addressing the difficult issue of testing behaviour or dynamic knowledge. Our approach attempts to combine unit testing techniques with the adoption of test cases in Test Driven Development (TDD) (Janzen & Saiedian, 2005). This is a useful hybrid approach for addressing the complex interactions of task behaviour and objects. The general intuition adopted from the TDD approach of testing is that it should be "done early and done often". In TDD, a test case is written first and then the actual module is written. Writing test cases first can be beneficial because instead of thinking of test cases as "how do I break something", writing test cases first make you think "what do I want the program to do". In other words, writing test cases first make you think about what functionality and objects are required. Ontology development could also benefit from this kind of approach.

2.1.4 Lack of support for ontology visualization

Most ontology editing tools lack visualization support. Often ontology visualization tools use 2 dimensional on 2D graphics for representing concepts and relationships among concepts. 2D graphics are often employed for representing static knowledge in hierarchical or frame like structures. However, 2D graphics are not adequate for representing complex and related information because there is insufficient space on a bi-dimensional plane, where complex and related visual objects tend to squeeze together or even overlap with each other. In an effort to accommodate a model of complex related objects in a 2D plane, existing visualization tools either collapse lower priority visual objects into high level objects, or implement different representations that offer multi-perspectives on an organization of knowledge elements. While these approaches solve the problem of related objects being overcrowded in a shared space, it sacrifices clarity of the display.

A popular ontology editor is Protégé, which has some degree of visualization capabilities in that it generates a tree view of classes. The Protégé's ontology visualization applications are implemented in its plug-ins; some examples of these ontology visualization plug-ins include Jambalaya (Wu & Storey, 2000) and OntoSphere (Bosca et al., 2005), which are discussed as follows.

Jambalaya is an ontology visualization tool plug-in in Protégé. It provides several viewing perspectives for the ontology model, thereby enhancing user browsing, exploring and interacting with a 2D ontology visualization. Jambalaya only visualizes the static knowledge of classes and instances of an application ontology, it does not support dynamic knowledge visualization. Furthermore, since Jambalaya is based on 2D graphics, the space it supports is insufficient for rendering complex knowledge. In its representation, text labels and symbols tend to overlap when the domain ontology is represented as a hierarchy involving many levels of concepts. This deficiency means it is difficult for users to view and understand the concepts and the relationships among concepts when the domain ontology is complex.

OntoSphere is also a Protégé plug-in. By employing 3D graphics for visualization, OntoSphere extends the volume of space available for visualizing overcrowded concepts. The 3-dimentional or 3D view is natural for humans. And its main advantage is that it allows users to manipulate the visualized knowledge elements of the application ontology by means of the actions of zooming, rotating and translating. Through physical manipulation of the concepts, the user can better understand a complex ontology. For this purpose, OntoSphere provides four scenes so that the user can observe a visualized application ontology from multiple perspectives.

However, the weakness of OntoSphere include (1) it was not developed based on any ontological engineering methodology, and (2) it does not support visualization of dynamic knowledge. Although the employment of 3D graphics enlarges the space available for rendering images, the problem of overlapping concepts and labels still exists when the application ontology is complex.

2.1.5 Lack of support for ontology management

Currently ontology development within the SW is not managed in a systematic manner. Anyone can create, reuse, and/or extend concepts in a distributed and heterogeneous environment, and different versions of an ontology can be created, which results in backwards compatibility problems. A system for ontology management is needed in order to document, track, and distribute ontologies in such an environment. Existing ontology

management frameworks provide inadequate support for replicating an ontology within the public domain and detecting when a public domain ontology has been tampered with. In addition, weaknesses of existing ontology management systems include vulnerability to malicious ontology authors, little consideration for intellectual property rights, and lack of support for ontology versioning.

2.1.6 Specific research objectives

In view of the five weaknesses discussed above, the overall objective of our research work is to construct a suite of software tools for modeling, visualizing and managing knowledge. The tools are designed so that they: (1) are derived from a theoretical basis of a knowledge modeling technique, (2) can sufficiently express the diverse types of knowledge required to model an ontology of an industrial domain, which include both static and dynamic knowledge, (3) can support visualization of both static and dynamic knowledge, (4) can support ontology management, and (5) can support testing of the developed knowledge model. In this chapter, the tools of Onto3DViz and Distributed Framework for Knowledge Evolution (DFKE) (Obst, 2006) are discussed which present new approaches for supporting ontology visualization and management. Dyna is a tool that has been developed for tackling dynamic knowledge modeling and ontology testing, and it is presented in detail in (Harrison & Chan, 2009). The two tools that will be discussed here have been developed based on the knowledge modeling method of Inferential Modeling Technique (IMT). Since the tools are based on the weak theory provided by the IMT, this knowledge modeling technique is presented next.

2.2 Inferential modeling technique

The Inferential Modeling Technique is a knowledge engineering method that supports developing a domain ontology consisting of both static and dynamic knowledge of the problem domain. Static knowledge consists of concepts, attributes, individuals and the relationships among them; dynamic knowledge includes objectives, tasks, and relationships among objectives and tasks. Static and dynamic knowledge are intertwined in that a task is a process that manipulates static knowledge to achieve an objective. The details of this modeling technique can be found in (Chan, 2004).

3. Conclusion design and implementation of the tools

The two tools that were designed and developed to address the weaknesses of ontology visualization and management include Onto3DViz and DFKE (Obst, 2006). Onto3DViz is an ontology visualization tool developed based on the IMT, which supports visualization of both static and dynamic knowledge models. Onto3DViz also supports knowledge sharing and re-use, by requiring ontology documents represented in OWL and XML as the input. DFKE is an ontology management tool which incorporates a monotonic versioning system with a peer-to-peer distributed knowledge management system. This tool was designed and a prototype of the system was developed.

3.1 Design of Onto3DViz

Onto3DViz can be seen as an extension of Dyna (Harrison & Chan, 2009) developed at the Energy Informatics Laboratory. Dyna is a dynamic knowledge modeler developed based on

the IMT, which can model the dynamic knowledge of an industrial domain. Dyna can be used in conjunction with Protégé, which models the static knowledge. When the two tools are used together, they can represent the static and dynamic knowledge of most industrial domains.

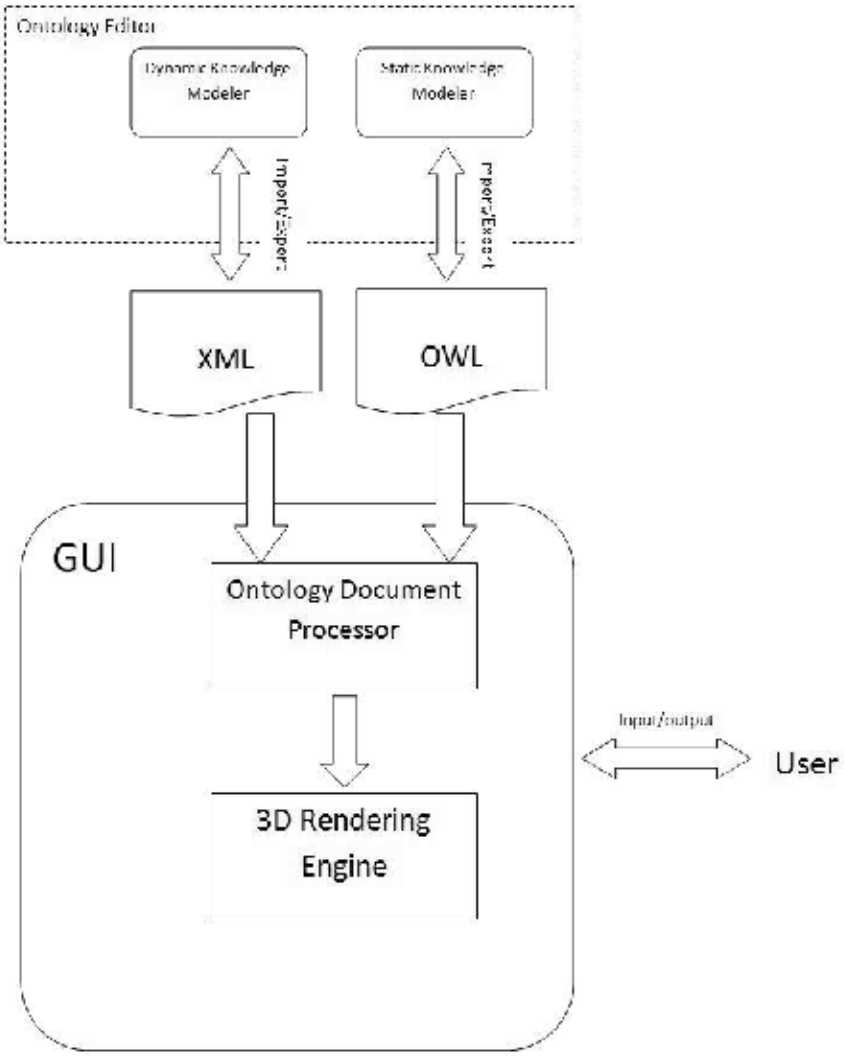


Fig. 1. High-level architecture of Onto3DViz

While Dyna (Harrison & Chan, 2009) can support ontology modeling, its capability for supporting ontology visualization is limited. Onto3DViz addresses this limitation and has been developed for visualizing an application ontology in 3D graphics. It is written in the Java™ language and its 3D visualization engine is implemented in Java 3D™. Onto3DViz differs from other ontology visualization tools in two ways: (1) it is a visualization tool developed based on the IMT (Chan, 2004), and (2) it supports visualization of both static and dynamic knowledge models specified according to the IMT. As well, Onto3DViz requires supports knowledge sharing and re-use. Unlike Dyna which has been implemented as a

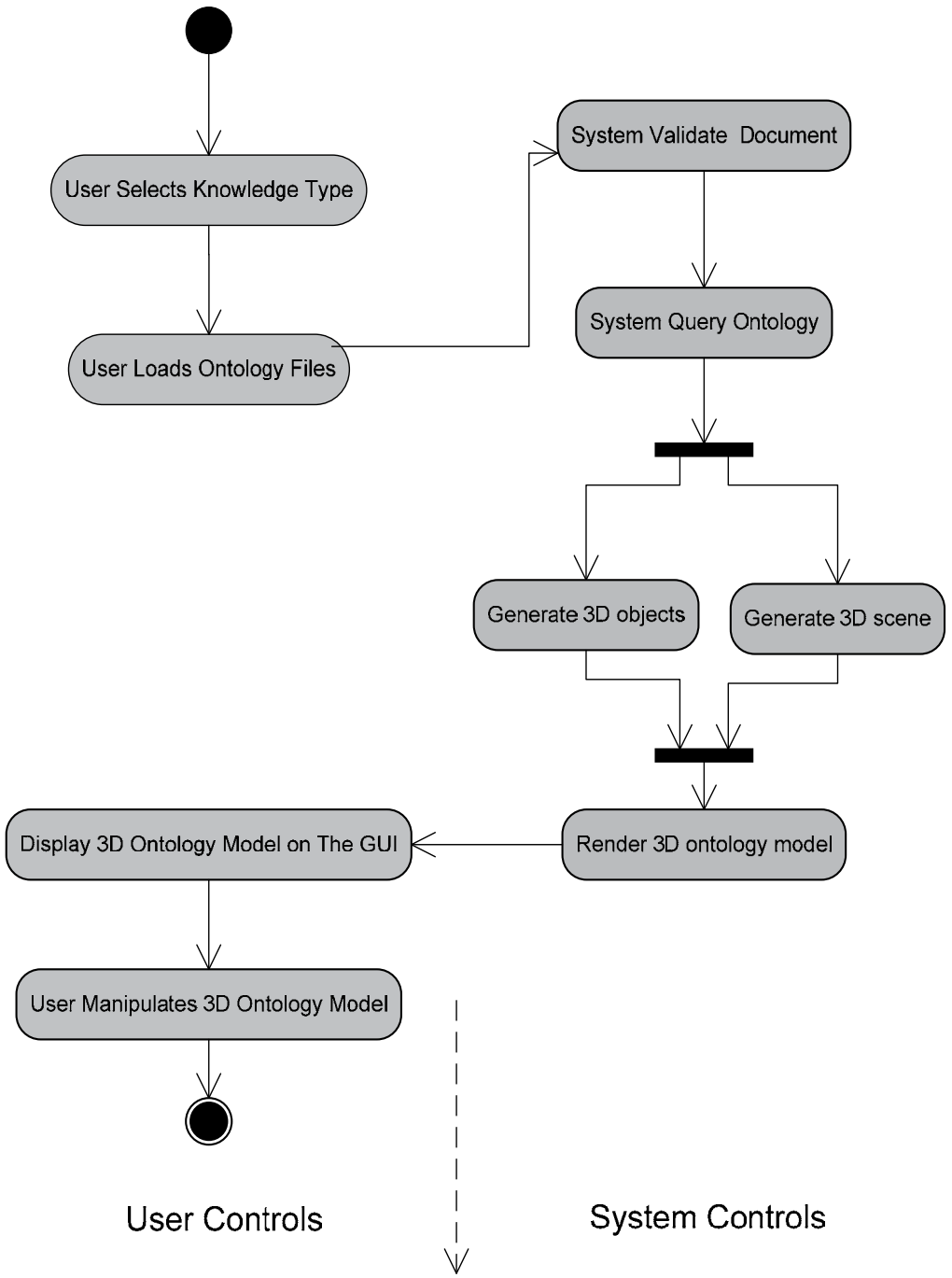


Fig. 2. Activity Diagram of Onto3DViz

Protégé plug-in, Onto3DViz is developed as an independent software application. It requires as input application ontologies stored in the OWL format, which can be developed based on IMT and by other tools or techniques. The ontology documents can be shared and re-used by other systems.

The high-level architecture of Onto3DViz is shown in figure 1. With the ontology editor, the user creates static knowledge models which consist of classes, properties, and relations, and dynamic knowledge models which consist of tasks, objectives, and behaviours of the tasks and links to the class objects. Then the files consisting of the XML and OWL represented ontologies are loaded into Onto3DViz, which generates the 3D visualized ontology model on the computer screen as an output. The user can then adjust the 3D model using the computer input devices of mouse and keyboard for exploring and navigating the visualized model.

The procedure of using Onto3DViz for ontology visualization involves nine steps. This process is also shown in the activity diagram in Figure 2.

1. The user selects the type of knowledge to be visualized, it can be either static knowledge only, or both static and dynamic knowledge.
2. The user selects an ontology document for loading into Onto3DViz.
3. Onto3DViz validates the OWL or XML file. If there is error, Onto3DViz will generate an error message to the user.
4. Onto3DViz extracts the ontology model using the OWL or XML parser and pass the retrieved information to the 3D graphic engine.
5. Onto3DViz generates the 3D ontology models.
6. The 3D graphic engine generates visual objects and visual effects.
7. The GUI displays the ontology 3D visualization model on the screen.
8. The user can adjust the viewing of the 3D visualization model by using the computer keyboard or mouse.
9. Onto3DViz provides adjusted views according to the user's preferences.

The modeling process assumes the user to have no programming knowledge. The tool is designed so that when the user loads a compatible and valid ontology document to Onto3DViz, a visualized ontology model will be constructed and displayed in 3D on the computer screen.

3.1.1 Knowledge extraction and interpolation

After an application ontology has been constructed using a tool like Dyna (Harrison & Chan, 2009) both the static and dynamic knowledge elements are stored in the XML files, which can be shared and re-used by other systems. Instead of being connected to an ontology editor that generates an application ontology, Onto3DViz is a separate stand-alone system that uses OWL and XML documents as inputs for 3D ontology visualization. The benefit of this approach is that Onto3DViz can produce a 3D ontology visualization as long as it receives a valid ontology document from another system or via the network. Since the formats of OWL and XML are well standardized and recognizable, the only requirement for using Onto3DViz is that a valid ontology document is available to the system in either format. Onto3DViz processes a valid OWL or XML formatted ontology file as follows:

1. Import Ontology
    a. If the ontology includes only static knowledge, the user provides a correct path of the ontology file to Onto3DViz so that it can import it.

   b. If the ontology includes both static and dynamic knowledge, the user provides the paths to both the static and dynamic knowledge files to Onto3DViz so that it can import the ontology from both paths.
2. Extraction and Interpolation of Knowledge
   a. Onto3DViz extracts and interpolates the OWL file(s) by applying a Protégé-OWL API. The Protégé-OWL API originally is one of the plug-in packages from the Protégé-OWL editor. This API is efficient in extracting and interpolating knowledge from an OWL file.
   b. Onto3DViz parses an XML-formatted file that stores dynamic knowledge by using Document Object Model (DOM[4]).
3. After the knowledge has been extracted from the ontology, it is stored into the Java objects that are custom designed for storing visual data in the computer memory, and which are used by Onto3DViz for creating the 3D model.

3.1.2 Visual object creation and display

The objective of Onto3DViz is to visually represent knowledge elements specified in the IMT. The tool employs three techniques for accomplishing this objective; they are (1) representation of concepts by the shape of visual objects, (2) size scaling of visual objects, and (3) positioning of visual objects. These techniques enable Onto3DViz to represent an ontology with minimal literal description while enhancing visual effects of the representation. The three techniques are discussed in detail as follows.

3.1.3 Shape of visual objects

Onto3DViz uses the following symbols to represent the following concepts:

- Class : Sphere 
- Objective : Cylinder 
- Instance : Box 
- Task : Cone 
- Relationship between concepts : Line

3.1.4 Size scaling of visual objects

For the purpose of showing hierarchical structure among concepts of an ontology, Onto3DViz decreases the size of the visual objects that occupy the lower levels in the hierarchy. For example, if the class of Thing is the parent class of all classes, then it is scaled to be the largest sized object. A subclass of the Thing class is a class, and the visual object of this class is scaled to be 70 percent of the size of the Thing class. Similarly, subsequent children classes are scaled to be 70 percent of the associated parent class. Hence the classes in the default hierarchical view of an application ontology represents the visual objects that are decreasing in size from top to bottom. By employing this technique, the hierarchical relationships as reflected in successive parent-and-child relationships are clearly represented among concepts.

[4] Document Object Model, http://www.w3.org/DOM/

3.1.5 Positioning of visual objects

Every 3D system requires a set of 3D coordinates to define its position and orientation. The 3D coordinate system used in Onto3DViz is defined as shown in figure 3.

This definition of the 3D coordinates in Onto3DViz is commonly adopted in most 3D computer graphic languages. The X axis is a horizontal line on a flat piece of paper, going from left to right, and the numbering goes from negative on the left to positive on the right. The Y axis is a vertical line on a piece of paper, and the numbering goes from negative in the bottom to positive at the top. The Z axis is perpendicular or orthogonal to the piece of paper, and the numbering follows the direction that is pointing out from the surface of the piece of paper. The intersection of these three axes is the origin. Onto3DViz allocates visual objects into three X-Y planes along the Z axis. They are:

- Class plane: this is the first X-Y plane located at the positive region of the Z axis as shown in figure 4a. The visual objects of classes are placed on this plane.
- Instance plane: this is the second X-Y plane located at the origin of the Z axis as shown in figure 4b. The visual objects of instances are located on this plane. If there is no instance being created in the ontology, this plane is not visible in the 3D model.
- Objective and task plane: This X-Y plane is perpendicular to and located in the negative region of the Z axis as shown in figure 4c. This plane contains the dynamic knowledge elements. The objectives are placed at the top region of this vertical plane, and the tasks associated with the objectives are placed in the bottom region of this plane. If there is not any dynamic knowledge element in the ontology, this plane is not visible in the 3D model.

The technique of organizing the placement of knowledge concepts at multiple levels in the three dimensional space enables categorization according to the types of the knowledge concepts, but also solves the problem of concepts being overcrowded in a shared space.

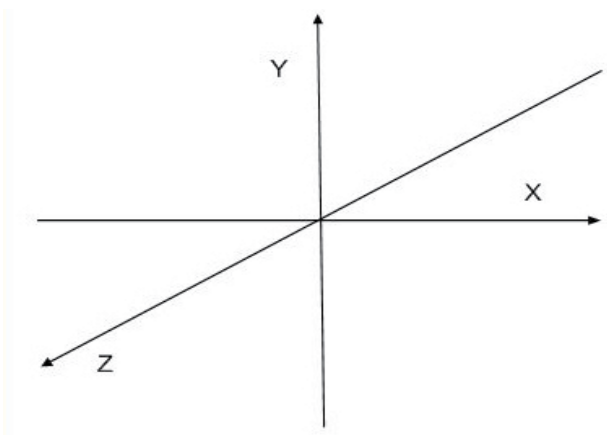


Fig. 3. The 3D coordinate system of Onto3DViz

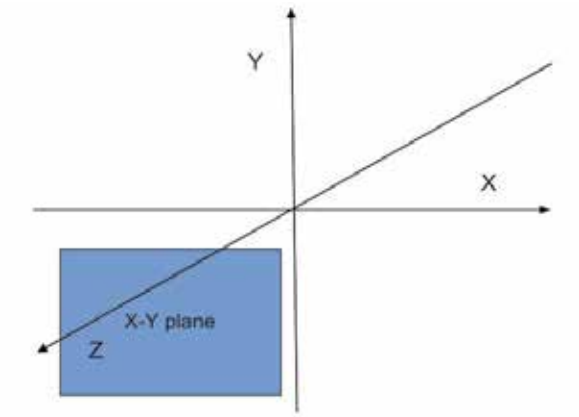


Fig. 4a. the X-Y plane stores classes

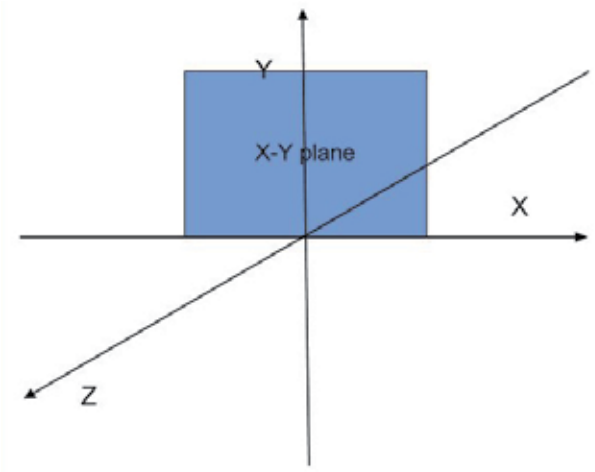


Fig. 4b. the X-Y plane stores instances

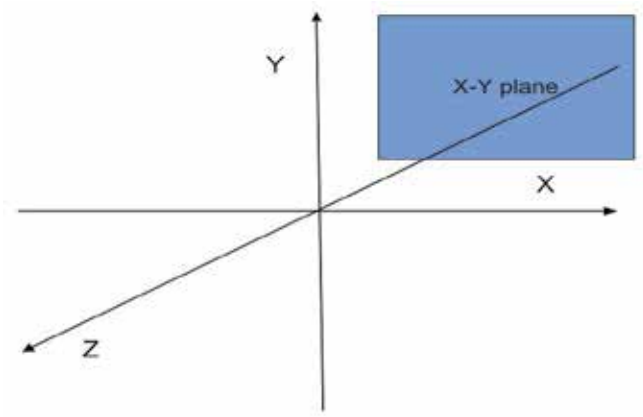


Fig. 4c. the X-Y plane stores objectives and tasks

3.1.6 User interaction

Interactions between the 3D model generated by Onto3DViz and the user are realized using the physical devices of the computer mouse and the keyboard. Both the mouse and keyboard can be used for controlling operation of the 3D model in Onto3DViz as follows.

1. Mouse:
   - Rotation of the model : the user can drag the mouse while pressing the left button
   - Translation of the model : the user can drag the mouse while pressing the right button
   - Zoom into the model : the user can drag the mouse while pressing the middle button (or holding the Alt key on the keyboard)
2. Keyboard:
   - Rotation
     - Rotate to the left : use ← Key
     - Rotate to the right : use → key
     - Rotate upward : use PageUp key
     - Rotate downward : use PageDown key
   - Translation
     - Translate along Z axis toward the front : use ↑ key
     - Translate along Z axis toward the back : use ↓ key
     - Translate along X axis toward the left : use ← key
     - Translate along X axis toward the right: use → key
     - Translate along Y axis upwards : use Alt-PageUp keys
     - Translate along Y axis downwards : use Alt-PageDown keys
   - Zoom
     - Zoom out : use – key
     - Zoom in : use + key
     - Reset to default viewing position : use = key

By combining these user control actions, the users can manipulate the model and obtain multiple perspectives of the 3D model of an application ontology.

3.2 Design of the ontology management system

To address the objective of knowledge management, a tool called Distributed Framework for Knowledge Evolution (DFKE) (Obst, 2006) has been developed. The high level architecture of DFKE is shown in figure 5.

The DFKE interacts with an ontology editor via the ClientServices. ClientServices provides functions for: (1) representing knowledge in a generic knowledge representation, and (2) importing and exporting knowledge models from and to XML. DFKE uses a peer-to-peer (P2P) network architecture. The "Feeder Node" is a node in this P2P network. The DFKE stores the knowledge projects in a knowledge repository or database, which is attached to a "Feeder Node". Network communication between the Client Services and the FeederNode (and Knowledge Repository) is achieved using the Common Object Request Broker Architecture (CORBA)[5]. DFKE also handles security, by encrypting network communication and digitally signing objects. There is also a Project Browser for assisting the user in finding projects on the P2P ontology network.

[5] CORBA, http://www.omg.org/gettingstarted/corbafaq.htm

One of the goals of the ontology management system was to hide as much of the internal processing of ontology management from the user as possible; the user should only be concerned with ontology modeling, not ontology management. With this goal in mind, much of the ontology management work is done automatically in the background for the user.Documentation is an important part of ontology management, which is often neglected. In creating ontologies for sharing and reuse, it is particularly important to provide information describing the author of the ontology. Author information includes the user's name, email address, organization, etc. Since author information is the same for all knowledge elements created by the user, the author information can be gathered once and then used by the system to automatically fill the input fields. The first time the user launches the Ontology Editor, a window requesting user information is displayed. This window has input fields for gathering personal identification information such as the user's name, email, web page, and organization. When a user creates any knowledge element, this gathered information is automatically attached to the knowledge element. All of this user information can be viewed in the property window of any knowledge element.

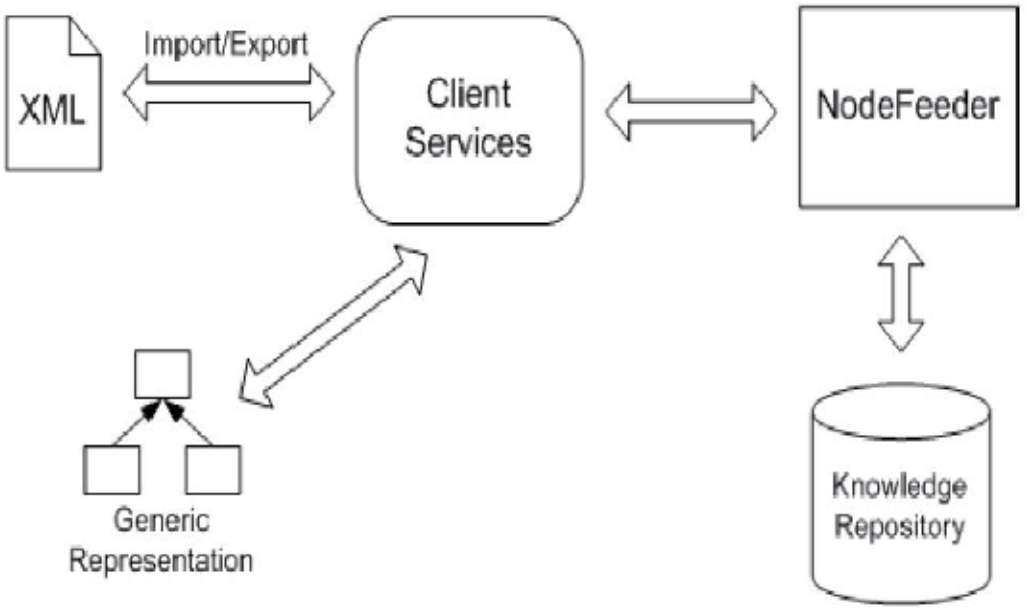


Fig. 5. High-level architecture of DFKE

The window also has an input field for a digital signature. A digital signature is "a data string which associates a message (in digital form) with some originating entity" (Maedche & Staab, 2003). The DFKE uses digital signatures to identify the author of an ontology element and to restrict access to an ontology element to only those people who should have access, thereby protecting the ontology from malicious ontology authors. To generate a digital signature for the user, the user enters a passphrase into an input field in the user information window, and clicks a button to generate the digital signature. In the background, the DFKE uses the user information and passphrase to generate the digital signature. This digital signature is used to identify each user and the knowledge elements the user creates. If another user attempts to modify a knowledge element that was created by the user, the other user will be denied.

The distributed and heterogeneous nature of the Semantic Web makes it very difficult for users to find ontologies. The Unified Ontology View (UOV) was proposed in (Harrison et al., 2005), and UOV provides a layer that hides the heterogeneous and distributed nature of the ontology network. The user is able to access the ontology network from anywhere and obtain a unified view of the ontologies in the network. Figure 9 gives an overview of the network environment which includes the servers, the ontology projects, and the client application. In this network, whether the Client is connected to server A, B, or C is irrelevant; the Client will be presented with the same view of the ontology projects in the ontology network.

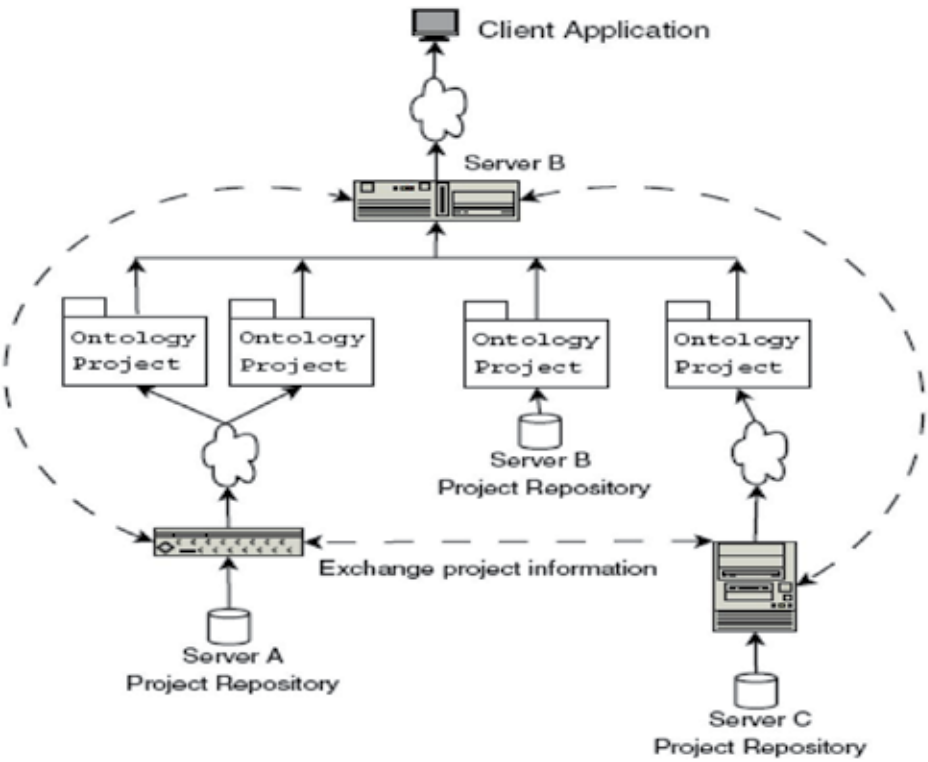


Fig. 6. Server view of ontology projects

Figure 7 shows the ontology projects from the perspective of the Client. When the Client connects to the ontology network, it only sees the ontology projects. It is possible that the ontology projects are distributed among many different types of systems, however, the Client does not need to be aware of this fact.

The ontology management system implemented the idea of a UOV by using a peer-to-peer (P2P) architecture, as shown in Figure 8. Knowledge projects created with the Ontology Editor are stored in a database or knowledge repository, which is connected to a Feeder node. The Feeder node that the Ontology Editor connects to may be local or remote. Feeder nodes can share knowledge projects with each other. The Feeder nodes discover other feeder nodes by contacting a central Hub node. Ontology projects are automatically added to the network when they are created with the Project Wizard. Once a project is added to a node on the network, it can be accessed from any node in the network.

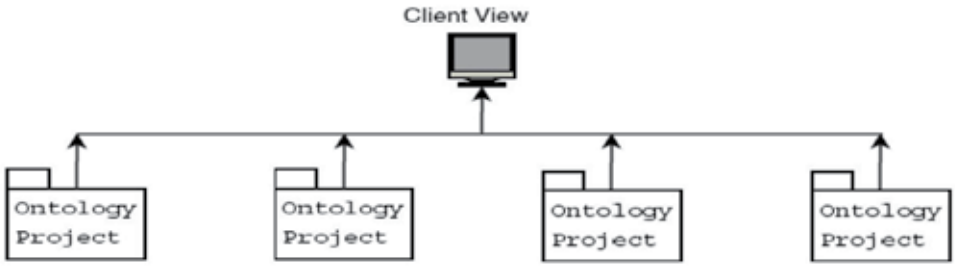


Fig. 7. Unified ontology view

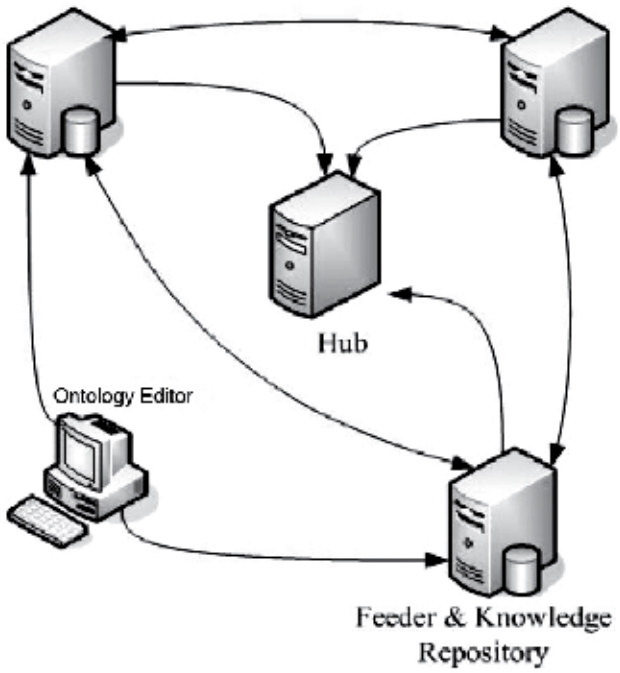


Fig. 8. P2P architecture

Projects can be used outside of the network by saving them in an XML file.
To support knowledge evolution, DFKE uses a generic ontology representation that was developed for representing conceptualizations and refinements of an ontology (Obst & Chan, 2005). Conceptualizations involve to be major changes to an ontology, while refinements reflect minor changes. The major/minor versioning scheme proposed by (Klein & Fensel, 2001) is used to keep track of the versions of the ontology as conceptualizations and refinements are added. Removing a conceptualization can have serious problems on other systems that depend on the conceptualization. Therefore, conceptualizations cannot be removed. The DFKE uses the concept of monotonicity to guarantee that existing concepts are not removed. The Class Editor helps support monotonicity by denying users the ability to remove knowledge elements that have been marked as monotonic by DFKE.

4. Discussions

Applications of the two tools have been conducted and presented in (Guo & Chan, 2010) and (Obst, 2009). Some strengths and weaknesses of the tools have been observed from the application experiences.

Onto3DViz is able to successfully render the complex domain concepts and relationships among concepts in a 3D model, thereby creating a visualized knowledge model that IMT formulated. The user can easily examine and manipulate the 3D model by performing the operations of zooming, rotating, and translating using the computer mouse or keyboard. The 3D model clearly shows the hierarchical structure of the static knowledge; it can also support representing the tasks associated with each object and arranging the tasks in the correct order. However, some weaknesses were also observed in the current version of Onto3DViz. First, when the visual objects are too close to each other, the objects and labels become overlapped, which makes it difficult for users to clearly see details of the model. Secondly, Onto3DViz does not currently support a search function. If the user wants to retrieve a particular concept, he or she has to manually examine the 3D model so as to find the corresponding visual object. This can be difficult if the ontology is complex or consists of a large number of concepts. Thirdly, if an application ontology consists of too many hierarchical levels, the lower level nodes will be small and difficult to see. This happens because the scaling technique implemented in Onto3DViz (described in section 3.1.4) reduces the size of the visual nodes in the lower levels.

The biggest limitation of the DFKE is the lack of integration with an ontology editor, such as Protégé and Dyna. This limitation means the user would need to conduct ontology modeling and ontology management as two different tasks and on two separate software tools. This is not convenient and DFKE in its current version does not adequately support the ontology authoring process.

5. Conclusions and future works

The objective of this research is to develop a suite of ontological engineering tools which supports (1) static and dynamic knowledge visualization, and (2) management of an application ontology. The motivation for this objective is derived from an analysis of existing tools which reveals their lack of support for modeling and visualizing dynamic knowledge. Among the few tools that support visualization, their capabilities for visualizing a large volume of information is limited due to the constraints of the 2D graphical space. Another weakness observed from our analysis is that existing ontology management frameworks provide inadequate support for replication and evolution of ontology, and they do not support detecting when a public domain ontology has been possibly tampered with. To address these limitations, Onto3DViz has been developed to support 3D visualization of an ontology model and the DFKE was designed for ontology management.

However, Onto3DViz and the monotonic DFKE have some weaknesses which will be tackled in the future. The current version of Onto3DViz can be enhanced by adding a collision avoidance system to address the problem of overlapping concepts. The collision avoidance system would verify that the visual objects in a 3D model do not collide. And if a collision is detected, it will automatically reassign the location of the visual objects and adjust the space between them. To address the difficulty of finding concepts in a model, a feature that supports automatic identification of a required concept can be added to Onto3DViz. To prevent overly tiny visual nodes from being generated, the next version of

Onto3DViz will set a cut off level for applying the scaling technique. That is, the system will include a pre-determined cut-off or optimal level for applying the scaling technique, so that the descendant nodes will remain the same size after the cut off level is reached. Moreover, assigning colors to the visual objects can also assist the user in indentifying different types of concepts. For instance, the visual objects for classes can be blue and visual objects for instances can be yellow. The different colors can help to distinguish the different knowledge types in a visualized ontology. Onto3DViz can also be implemented as a Protégé-OWL editor plug-in, so that ontology visualization and editing can be done in real time. As well, more user controls can be implemented which will enhance user-friendliness of the tool.
The DFKE has been designed and implemented as a standalone application and is not integrated with any other ontology tool as mentioned in section 4. This limits usefulness of the tool because ontology development and management have to be conducted separately. In the future, it will be useful to integrate DFKE with other ontological engineering tools, such as Protégé (Protégé), Dyna (Harrison & Chan, 2009), and Onto3DViz, so that users can conduct ontology authoring, editing, visualization and management in a unified system.

6. Acknowledgment

We are grateful for the generous support of Research Grants from Natural Sciences and Engineering Research Council (NSERC) and the Canada Research Chair Program to the first author.

7. References

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American* , 34-43.

Bosca, A., Bonino, D., & Pellegrino, P. (2005). OntoSphere: more than a 3D ontology visualization tool. *Proceedings of SWAP 2005, the 2nd Italian Semantic Web Workshop.* Trento, Italy: CEUR Workshop Proceedings.

Chan, C. (2004). From Knowledge Modeling to Ontology Construction. *Int. Journal of Software Engineering and Knowledge Engineering, 14(6)* .

Cranefield, S., Pan, J., & Purvis, M. (2005). "A UML Ontology and Derived Content Language for a Travel Booking Scenario", *Ontologies for Agents: Theory and Experiences* Basel: Birkh\auser (2005), p. 259-276.

Djuric, D., Gasevic, D. & Devedzic, V. (2005). "Ontology Modeling and MDA", In *Journal of Object Technology*, 4(1), 2005, pp. 109-128.

Fensel, D., Van Harmelen, F., Ding, Y., Klein, M., Akkermans, H., Broekstra, J,. . . Horrocks, I. (2002). "On-To-Knowledge in a Nutshell", 2002.

Gomez-Perez, A. Fernandez-Lopez, M., & Corcho, O. (2004). "Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce, and the Semantic Web", Springer, 2004.

Guo, S. & Chan, C. (2010). A tool for ontology visualization in 3D graphics: Onto3DViz. *The Proceedings of the 23rd Canadian Conference on Electrical and Computer Engineering*. Calgary, AB, May 2010.

Gruber, T. (1993). Towards Principles for the Design of Ontologies Used for Knowledge. *In Guarino & Poli (eds): Formal Ontology in Conceptual Analysis & Knowledge Representation* .

Harrison, R. & Chan, C. (2005). "Implementation of an Application Ontology: A Comparison of Two Tools", *Artificial Intelligence Applications and Innovations II: Second IFIP TC12 and WG12 Conference on Artificial Intelligence Applications and Innovations (AIAI-2005)*, Beijing, China, Sept. 7-9, 2005, pp. 131-143.

Harrison, R. Obst, D., & Chan, C. (2005). "Design of an Ontology Management Framework", In Proceedings of the 4th International Conference on Cognitive Informatics 2005 (ICCI'05), University of California, Irvine, USA, Aug. 8-10, 2005, pp 260-268.

Harrison, R. & Chan, C. (2009). "Dynamic Knowledge Modeler" *Artificial* Intelligence *for Engineering Design, Analysis and Manufacturing,* Volume 23 Issue 1, February 2009 Cambridge University Press New York, NY, USA.

Janzen, D. & Saiedian, H. (2005). "Test Driven Development: Concepts, Taxonomy, and Future Direction", IEEE Computer, September 2005.

KANO. (n.d.). Retrieved on Thursday 6 January 2011. < http://kaon.semanticweb.org/>

Obst, D. & Chan, C.W. (2005). "Towards a Framework for Ontology Evolution", *18th Annual Canadian Conference on Electrical and Computer Engineering (CCECE05)*, Saskatoon, Saskatchewan, Canada, May 1-4, 2005.

Obst, D. (2006). "Distributed Framework for Knowledge Evolution", Presentation at University of Regina Grad. Student Conf., Regina, SK, Canada, 2006.

Obst, D. (2009)., "Monotonic Versioning of Ontologies in a Peer-to-Peer Environment", Master Thesis, University of Regina, July 2009.

OntoBuilder. (n.d.). Retrieved on Thursday 6 January 2011. <http://iew3.technion.ac.il/OntoBuilder/>

Ontolingua. (n.d.) Retrieved on Thursday 6 January 2011. <http://www-ksl-svc.stanford.edu:5915/>

OWL Unit Test Framework. (n.d.). Retrieved on Thursday 6 January 2011. < http://www.co-ode.org/downloads/owlunittest/>

Protégé. (n.d.). Retrieved on Thursday 6 January 2011. http://protege.stanford.edu

Protégé OWLViz. (n.d.). Retrieved on Thursday 6 January 2011. <http://www.co-ode.org/downloads/owlviz/co-ode-index.php>

Klein, M. & Fensel, D. (2001). "Ontology Versioning on the Semantic Web", *Proceedings of SWWS '01, The First Semantic Web Working Symposium*, pp. 75-92, August 2001.

Maedche, A. & Staab, S. (2003). "Ontology Learning", In: Staab, S., Studer, R., eds Handbook of Ontologies in Information Systems, Springer Verlag, pp 173-190, 2003.

McGuiness, D., Fikes, R., Rice, J., & Wilder, S. (2000). "An Environment for Merging and Testing Large Ontologies". In *Proceedings of KR 2000*, pp. 485-493.

WebODE. (n.d.). Retrieve from Thursday 6 January 2011. <http://webode.dia.fi.upm.es/WebODEWeb/index.html>

Wu, J. & Storey, M.-A. (2000). A Multi-Perspective Software Visualization Environment. *conference of the Centre for Advanced Studies on Collaborative Research 2000*, (pp. 31-40). Mississauga, Ontario, Canada.

4

Data Quality Enhancement Technology to Improve Decision Support

Ahmad Shahi, Rodziah binti Atan and Nasir bin Sulaiman
University Putra Malaysia
Malaysia

1. Introduction

Uncertainty is a very important aspect of real human life. It means "Not knowing with certainty; such as cannot be definitely forecast" [1]. This uncertainty in fuzzy systems occurs mainly due to volume of work, lack of theoretical knowledge and lack of experimental results [2].

Mendel [3] has noted that uncertainty also exists while building and using typical fuzzy logic systems. He has described four sources of uncertainty: *Uncertainty about the meanings of the words that are used in a rule*, this is the uncertainty with the membership functions because membership functions represent words in a fuzzy logic system. It can be both antecedents and consequents; *Uncertainty about the consequent that is used in a rule,* this is the uncertainty with the rule itself. A rule in FLS describes the impact of the antecedents on the consequent. Expert may vary in their opinion to decide this nature of impact; *Uncertainty about the measurements that activate the FLS,* this is the uncertainty with the crisp input values or measurements that activates the FLS. These measurements may be noisy or corrupted. The noise can again be in a certain range or totally uncertain meaning stationary or non-stationary; *Uncertainty about the data that are used to tune the parameters of a FLS*, this is the uncertainty with the measurements again.

To deal with the uncertainty, fuzzy logic is a proper way to model human thinking. Although it was introduced by Lotfi Zadeh in 1965, it has been used to build expert systems for handling ambiguities and vagueness associated with the real world problems which involve different kinds of uncertainty [4].Thus in order to strengthen fuzzy system model, quality of data as an input of the model should be enhanced. Outliers and noisy data, these uncertainty arise from mechanical faults, change in system behavior, fraudulent behavior, network intrusions, sensor and device error, human error and so on [5, 6]. However, to strengthen fuzzy system model, outliers should be isolated that, the following section demonstrates about details of isolating outliers.

1.1 The reason of isolating outliers

The main reason for isolating outliers is associated with data quality assurance. The exceptional values are more likely to be incorrect. According to the definition, given by Wand and Wang [7], unreliable data represents an unconformity between the state of the database and the state of the real world. For a variety of database applications, the amount of erroneous data may reach ten percent and even more [8]. Thus, removing or replacing

outliers can improve the quality of stored data. Isolating outliers may also have a positive impact on the results of data analysis and data mining. Simple statistical estimates, like sample mean and standard deviation can be significantly biased by individual outliers that are far away from the middle of the distribution. In regression models, the outliers can affect the estimated correlation coefficient [9]. Presence of outliers in training and testing data can bring about several difficulties for methods of decision-tree learning, described by Mitchell in [10] and parameters in Gaussian membership function parameters in [2]. For example, using an outlying value of a predicting nominal attribute can unnecessarily increase the number of decision tree branches associated with that attribute. In turn, this will lead to inaccurate calculation of attribute selection criterion (e.g., information gain). Consequently, the predicting accuracy of the resulting decision tree may be decreased. As emphasized in [11], isolating outliers is an important step in preparing a data set for any kind of data analysis.

1.2 Effective quality of data on technology of fuzzy system

Fuzzy systems are expressed by membership functions. The outlier and noise are kinds of uncertainty which have effect on the membership function parameters, such as the Gaussian membership. In Gaussian, there are two parameters, mean and standard deviation, which are tuned based on the dataset. However, if the desired data is extracted from the dataset, Mean and Standard deviation can be accurate parameters for the Gaussian membership. Hence, to make a robust model, the outliers must be detected and the noisy data must be removed from the dataset.

There is a direct, although rarely explored, relation between uncertainty of input data and fuzziness expressed by Membership Functions (MFs). Various assumptions about the type of input uncertainty distributions change the discontinuous mappings provided by crisp logic systems into more smooth mappings that are implemented in a natural way by fuzzy rules using specific types of MFs. On the other hand shifting uncertainty from fuzzy rules to the input values may simplify logical rules, making the whole system easier to understand, and allowing for easy control of the degree of fuzziness in the system [12].

If regions of the data of different classes are highly overlapping or if the data is noisy, the values of the membership degrees could be misleading with respect to rule confidence if the core region is modeled too small. In fact, we show that data regions with a high membership degree need not to be the regions with a high rule confidence. This effect that we call membership is unrobustness [2].

Therefore, the Fuzzy C-Mean clustering (FCM) is utilized to detect the outlier and statistic equation is used to remove the noisy data in order to improve the quality of the data.

2. Design of method

As shown in Figure 1, the Weka software which was developed at Waikato University [13], is used for pre-processing the data in the dataset. After cleaning the data, the FCM with statistic equation (which is described in following section) were utilized to detect outliers, remove noisy data and extract the desired data to get data of high quality. In the next step, Type-1 FLS with gradient descent algorithm) were used to make a decision on such data, after analyzing the data to decide on the parameters to be used, including temperature, humidity and so on. The important part of this technique is that the gradient descent

algorithm was used to tune the membership function parameters. The proposed method can be seen in Figure 1. The details of the proposed method are described in following sections.

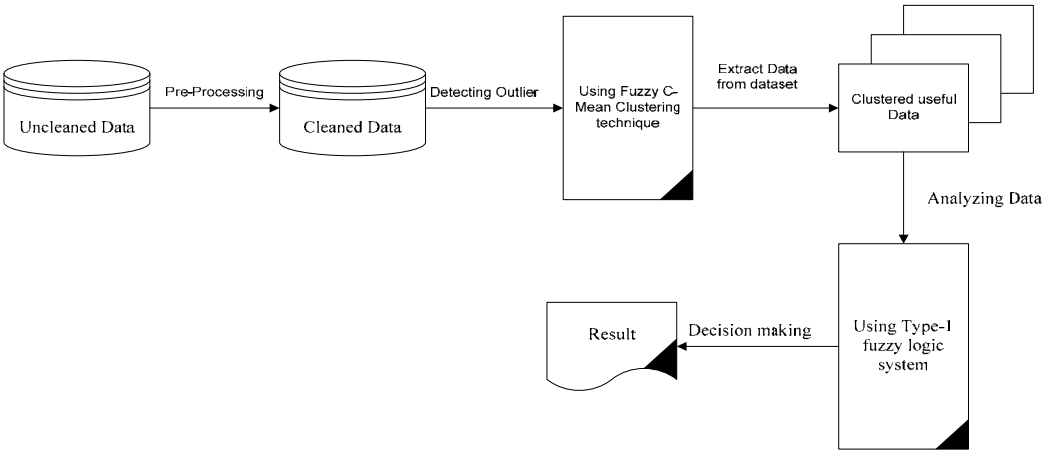


Fig. 1. Flowchart of Proposed method

2.1 Fuzzy C-Mean clustering

In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to one cluster. To manage uncertainty in getting data quality and accurate model, a new method known as the Fuzzy C-Mean (FCM) clustering is needed.

The FCM computes the distance measured between two vectors, where each component is a trajectory (function) instead of a real number. Thus, this Fuzzy C-mean clustering is rather flexible, moveable, creatable, as well as able to eliminate classes and any of their combination. From the huge number of clustering methods, the fuzzy clustering was focused on in the methodology of the present study since the degree of membership function on an object to the classes found provides a strong tool for the identification of changing class structures. Thus, the Fuzzy C-Means is used in order to build an initial classifier and to update the classifier in each cycle; nevertheless, the methodology presented can still be extended to any other techniques which determine such degrees of membership (e.g. probabilistic clustering, etc.) [14].

Before the FCM could be utilized, the noise was removed from the dataset due to affect on clustering data. Based on the statistic definition:

According to [15] a noise is considered to be more than three standard deviations away from the mean which is formulated as below:

$$\text{Noises} = \text{abs}\,(\text{object} - \text{MeanMat}) > 3*\text{SigmaMat};$$

In which, Meanmat is mean, SigmaMat is Standard deviation and abs is the functioning in mathematic for Absolute value, i.e. instance: abs (-5) =5, which was implemented in MATLAB software.

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. The method developed by [16, 17] is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \ , 1 \le m < \infty \tag{1}$$

Where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension centre of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the centre.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centre c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}}, \tag{2}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{3}$$

This iteration will stop when the $\max_{ij}\left\{\left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, and k is the iteration step. This procedure converges to a local minimum or a saddle point of J_m.

The algorithm is composed of the following steps [17]:

1. *Initialize* $U=[u_{ij}]$ *matrix,* $U^{(0)}$
2. *At k-step: calculate the centre vectors* $C^{(k)}=[c_j]$ *with* $U^{(k)}$

   $$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m}$$
3. *Update* $U^{(k)}$, $U^{(k+1)}$

   $$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}}$$
4. *If* $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$ *then STOP; otherwise return to step 2.*

The Fuzzy C-Mean clustering and statistic equation were implemented in the MATLAB software.

2.2 Type-1 fuzzy logic system

Fuzzy logic was developed by Lotfi Zadeh a professor at the University of California, Berkley. Fuzzy system is useful for real world problems where there are different kinds of uncertainty [18]. The idea of fuzzy logic was to show that there is a world behind conventional logic. This kind of logic is the proper way to model human thinking. Fuzzy logic is recently getting the attention of artificial intelligence researchers. It is being used to build expert systems for handling ambiguities and vagueness associated with real world problems. Figure 2 shows the architecture of Type-1 Fuzzy system with gradient descent algorithm.

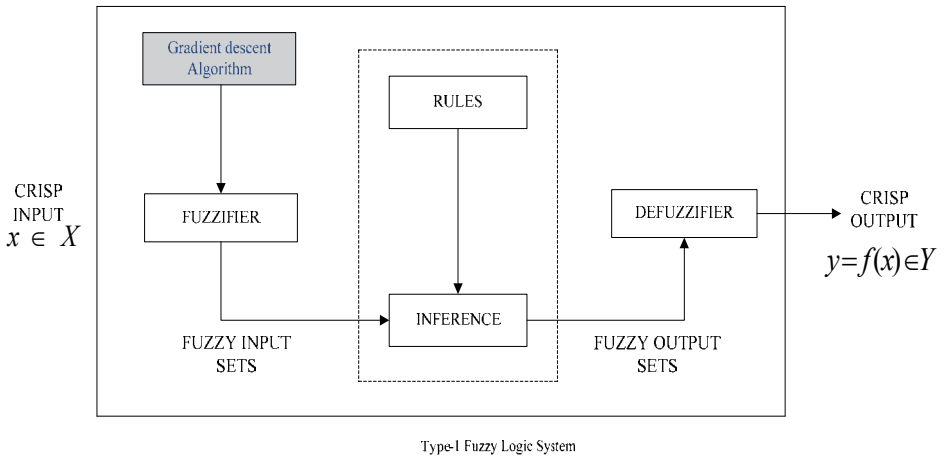


Fig. 2. Structure of Type-1 Fuzzy Logic system with Gradient Descent Algorithm

Gradient descent Algorithm: Gradient descent technique is a training algorithm which was used to tune the membership function parameters in fuzzy system. Using this algorithm the membership function parameters can be optimized and the error rate is reduced to get more accurate results. The details of the algorithm were explained in [19].

Fuzzification Process: According to [20] fuzzifying has two meanings. The first is the process fining the fuzzy value of a crisp one. The second is finding the grade of membership of a linguistic value of a linguistic variable corresponding to a fuzzy or scalar input. The most used meaning is the second. Fuzzification is done by membership functions.

Inference Process: The next step is the inference process which involves deriving conclusions from existing data [20]. The inference process defines a mapping from input fuzzy sets into output fuzzy sets. It determines the degree to which the antecedent is satisfied for each rule. These results in one fuzzy set assigned to each output variable for each rule. MIN is an inference method. According to [21] MIN assigns the minimum of antecedent terms to the matching degree of the rule. Then fuzzy sets that represent the output of each rule are combined to form a single fuzzy set. The composition is done by applying MAX which corresponds to applying fuzzy logic OR, or SUM composition methods [20].

Defuzzification Process: Defuzzification is the process of converting fuzzy output sets to crisp values [20]. According to [22] there are three defuzzification methods used : *Centroid*, *Average Maximum* and *Weighted Average*. Centroid method of Defuzzification is the most commonly used method. Using this method the defuzzified value is defined by:

$$\text{Centroid}=\frac{\int x\mu(x)dx}{\int \mu(x)dx} \tag{4}$$

Where $\mu(x)$ is the aggregated output member function. The details of Fuzzy system have been explained in [20] [23].

3. Experiments and results

The dataset for preprocessing is taken from Politecnico in Torino (Italy) that contains two attributes, temperature and humidity. These measures have been taken from a sensor and recorded by a computer at regular intervals of about 15 minutes (the average interval was estimated at 896 seconds) [24]. The dataset has been preprocessed by WEKA software to check missing value. And the statistic equation was utilized to remove noisy data, after that FCM clustering was used to detect outliers. By detecting outliers, desired data was extracted from dataset as an input of fuzzy system for controlling and decision making.
The Italy dataset possesses 4600 instances. Based on the equation coded in the Matlab, the noise was found and removed from the dataset. The graphic view of the Italy dataset before removing noise is depicted in Figure below.

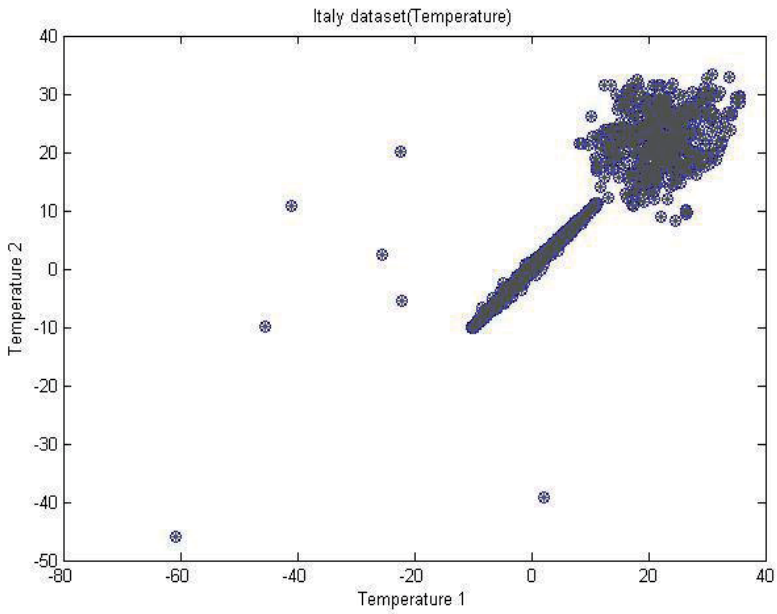


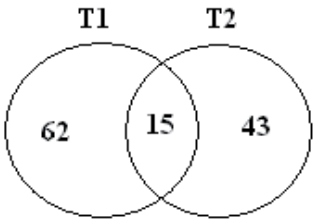
Fig. 3. Graphical View of the Original Italy Dataset

As shown in Figure 3, the attributes (temperature 1 and 2) were derived from the device sensor. The extracted data may include noisy data due to device and measurement errors. This would decrease the quality of the data. Therefore, the program in this study was applied based on the definition of the statistic equation on the data to remove the noisy data. The characteristics and results are shown in Table 1.

Attribute	Number of instances with noise	Number of noisy instances	Number of common noisy instances	Total noise	Number of instances without noise
Temperature 1	4600	62	15	90	4510
Temperature 2	4600	43			4510

Table 1. The Number of the Italy dataset instances with and without noise

Table 1 shows that the number of the noisy data in the attribute Temperature 1 is 62 and this is 43 for the attribute Temperature 2; 15 noise instances are common among the attributes Temperature 1 and Temperature 2, and thus the total instances for two attributes after removing noise is 4510. It means:



4600-(62+43)-15=4510

After removing the noise from the dataset, the Fuzzy C-Mean clustering (FCM) was used to detect the outliers and extract the desired data.

The FCM was used to cluster the data after removing the noisy data to mine the desired cluster. The noise detected in the present study was found to have effect on the FCM, and the mean is not defined well due to the noise. Therefore, the noisy data must be removed in order to get the accurate results.

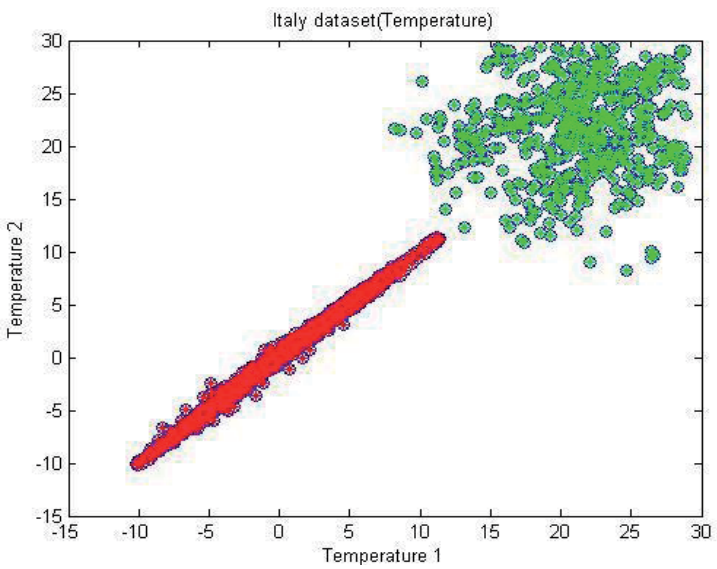


Fig. 4. Detecting and Clustering Outliers on Italy Dataset.

The attributes, which were derived after removing noise and clustered based on the FCM algorithm, are shown in Figure 4. The number of clusters is 2 (k=2, k is a number of cluster); green represents one cluster and red is another cluster which possesses behaviour that is different from each other.

Dataset	Number of cluster	Number of Instances with noise	Number of noisy instance	Number of instance without noise	Number of instances		Center of Cluster (Mean)	
					Cluster 1	Cluster 2	Cluster 1 (x,y)	Cluster 2 (x,y)
Italy	2	4600	90	4510	566	3944	(20.9173, 21.4507)	(0.7194, 0.7179)

Table 2. Summary of characteristics the Italy dataset

As shown in Table 2, are characteristics of Italy dataset. The dataset has 2 clusters; cluster 1 contains 566 records and cluster 2 has 3944 records. Based on the two dimensions, the centre of cluster 1: Temperature1=20.9173 and Temperature2=21.4507; likewise for cluster 2: Temperature1=0.7194 and Temperature2=0.7179.

3.1 Analysis using type-1 FLS

The dataset was pre-processed using the Weka to check for any missing values. After that, the data was used as an input for the method (i.e. Type-1 Fuzzy Logic System).

In Figure 5, blue lines show the release desired or the real data and red lines are obtained from the data predicted as the outcome of the method (Type-1 FLS).

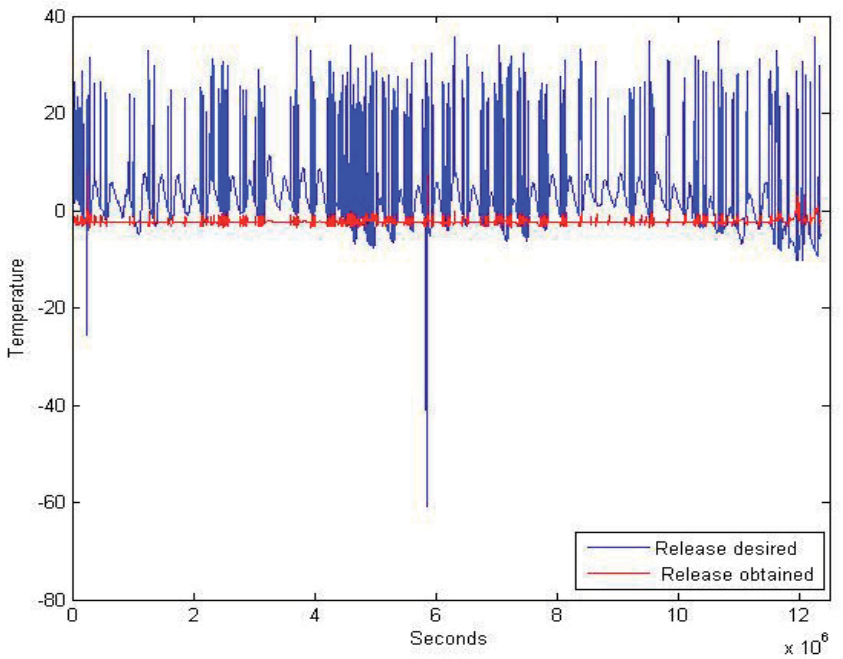


Fig. 5. Type-1 FLS on Italy dataset

Based on Figure 5, the release desired colored in blue (real data) and release obtained colored in red (predicted data) are not close to each other. The difference is due to the low quality of the data that contains noise and outliers, affecting the membership function and causing the function to be inaccurate and not robust. The membership function has two parameters, namely the mean and standard deviation, which are tuned based on the data.

3.2 Analysis and results using proposed method

After removing the missing value, FCM with statistic equation are applied to detect outliers and remove noisy data on the Italy dataset, the desired clustered data was extracted and entered as an input entry into the Type-1 Fuzzy Logic System with gradient descent algorithm to tune membership function parameters. The result shows a different type of graph that presenting the method effect to the dataset.

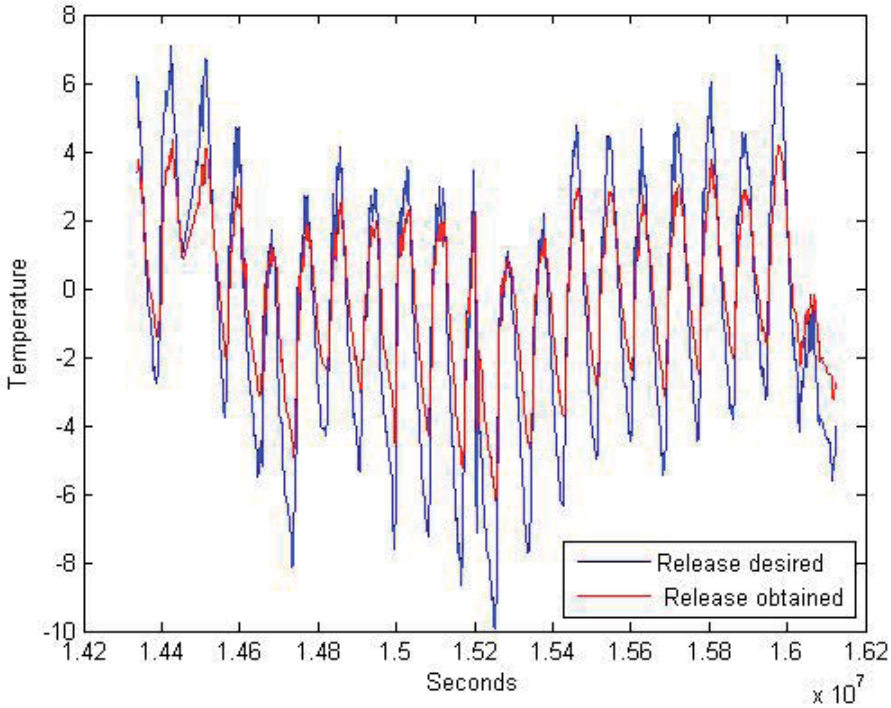


Fig. 6. Proposed Method on Italy Dataset

Figure 6 shows the output of Proposed. The attribute is a temperature based on time (second). A small scale of data (desired clustered data) was used to show the system behaviour (Proposed Method).

Blue shows the desired data or the real data and red is the obtained or predicted data. If data before the contribution (Type-1 FLS) were compared, it resulted much better and represents predicted data more closely to the real ones.

3.3 Accuracy measurements

To evaluate the results we used standard measurement called Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) which formulated as below:

If O is the actual observation for time period t and P is the forecast for the same period, then the error is defined as:

Error: $$e_t = O_t - P_t \tag{5}$$

Since there are observations and predictions for n time periods, then there will be n error terms, and the following standard statistical measures can be defined:

Mean Absolute Error: $$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{6}$$

Root Mean Square Error (RMSE): $$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2} \tag{7}$$

The MAE is defined by first making each error positive by taking its absolute value and then averaging the results. The RMSE is a quadratic scoring rule which measures the average magnitude of the error. The equation for the RMSE is given in both of the references. Expressing the formula in words, the difference between forecast and corresponding observed values are each squared and then averaged over the sample. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable.

The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE; the greater difference between them, the greater the variance in the individual errors in the sample. If the RMSE=MAE, then all the errors are of the same magnitude. Both the MAE and RMSE can range from 0 to ∞. They are negatively-oriented scores: Lower values are better. Each of these statistics deals with measures of accuracy whose size depends on the scale of the data [25].

Dataset: Italy (Temperature)	Before Contribution	After Contribution
Measurement metrics/methods	Type-1 FLS	Proposed Method
MAE	7.0321	1.3994
RMSE	10.5019	1.6590

Table 3. Accuracy Measurements of Italy Dataset

In such circumstances, Table 3 shows results of error value for type-1 FLS before and after contribution .For MAE measurement is 7.0321 and 1.3994 using type-1 FLS before and after clustering respectively. These values in RMSE measurement are 10.5019 and 1.6590. The result shows the method after contribution achieve minimum error rate in both MAE and RMSE. RMSE

These effective results are due to the mechanism of Type-1 FLS* with FCM* method. Type-1 FLS+ improves the quality of data by detecting outliers, removing noisy data and tuning MFs parameters by training algorithm is called Gradient Descent algorithm. Thus this method is a significant technique that can improve the accuracy of weather situation.

4. Conclusions

Outlier and noise are part of uncertainty that arises due to mechanical faults, changes in system behavior, fraudulent behavior, network intrusions, human errors, keyboard error, hand writing error and so on that affect on measurement of Gaussian membership function parameters. In Gaussian there are two parameters, Mean and Standard deviation that are tuned based on dataset, therefore if we do not extract useful knowledge or desired clustered data from dataset, Mean and Standard deviation will not be accurate parameters for Gaussian membership function. From the huge number of clustering methods, Fuzzy C-Mean clustering is flexible, moveable, creatable, elimination of classes and any their combination. Since the degree of membership function on an object to the classes found provides a strong tool for the identification of changing class structures. Fuzzy C-Mean in order to build an initial classifier and to update our classifier in each cycle, thus we utilized Fuzzy c-mean clustering with statistic equation to remove noisy data and detect outlier and mine valuable data to get accurate result with Type-1 Fuzzy Logic Systems and gradient descent algorithm.

By applying proposed method, the quality of data has been improved (As shown in Table 3). The proposed method enhanced the data quality. Thus, by improving the quality of data, the accurate decision making will be achieved in decision support system.

5. References

[1] Russell, G., *Grossett:"Webster's New Dictionary and Thesaurus"*. Geddes and Grosset Ltd., New Lanark, Scotland, 1990.

[2] Rahman, A., *Handling imprecision and uncertainty in software quality models.* 2005.

[3] Mendel, J.M., *Uncertain rule-based fuzzy logic systems: introduction and new directions*. 2000: Prentice Hall.

[4] Zadeh, L.A., *The concept of a linguistic variable and its application to approximate reasoning.* Information sciences, 1975. 8(3): p. 199-249.

[5] Last, M. and A. Kandel. *Automated detection of outliers in real-world data*. 2001: Citeseer.

[6] Cherednichenko, S., *Outlier detection in clustering*. 2005.

[7] Wand, Y. and R.Y. Wang, *Anchoring data quality dimensions in ontological foundations.* 1996.

[8] Wang, R.Y., M.P. Reddy, and H.B. Kon, *Toward quality data: An attribute-based approach.* Decision Support Systems, 1995. 13(3-4): p. 349-372.

[9] Clarke, E. and T. Coladarci, *i> Elements of Statistical Reasoning.</i> New York*. 1999, Wiley.

[10] Mitchel, T., *Machine learning.* Machine Learning, 1997. 48(1).

[11] Pyle, D., *Data preparation for data mining*. 1999: Morgan Kaufmann Pub.

[12] Rehm, F., F. Klawonn, and R. Kruse, *A novel approach to noise clustering for outlier detection.* Soft Computing-A Fusion of Foundations, Methodologies and Applications, 2007. 11(5): p. 489-494.

[13] Kirkby, R. and E. Frank, *WEKA Explorer User Guide for Version 3-5-3.* University of Waikato,(June 2006), 2006.

[14] Crespo, F. and R. Weber, *A methodology for dynamic data mining based on fuzzy clustering.* Fuzzy Sets and Systems, 2005. 150(2): p. 267-284.

[15] Macfie, B.P. and P.M. Nufrio, *Applied statistics for public policy*. 2006: ME Sharpe Inc.

[16] Dunn, J.C., *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.* Cybernetics and Systems, 1973. 3(3): p. 32-57.

[17] Bezdek, J.C., *Pattern Recognition With Fuzzy Objective Function Algoritms, Plenum Press.* New York, 1981.
[18] Zadeh, L.A., *Outline of a new approach to the analysis of complex systems and decision processes.* IEEE Transactions on Systems, Man, and Cybernetics, 1973. 3: p. 28-44.
[19] Li-Xin, W., *A course in fuzzy systems and control.* prentice Hall, 1997.
[20] Siler, W. and J.J. Buckley, *Fuzzy expert systems and fuzzy reasoning*. 2005: Wiley-Interscience.
[21] Hwang, G.J. and S.T. Huang, *New environment for developing fuzzy expert systems.* J INF SCI ENG, 1999. 15(1): p. 53-69.
[22] *http://www.jimbrule.com/fuzzytutorial.html*. Last accessed 4 November 2006.
[23] *http://www.urc.bl.ac.yu/manuals/adv/fuzzy.* Lased accessed 3 November 2006.
[24] Mencattini, A., et al. *Local meteorological forecasting by type-2 fuzzy systems time series prediction*. 2005.
[25] Mountis, A. and G. Levermore. *Weather prediction for feedforward control working on the summer data*. 2005.

5

Fine-Grained Diagnostics of Ontologies with Assurance

Stefan Rass, Fadi Al Machot and Kyandoghere Kyamakya
Alpen-Adria Universität Klagenfurt
Austria

1. Introduction

Description logics (DL) is a class of logics for knowledge modeling, which are derived from semantic networks. Essentially they are to be understood as accessible fragments of predicate logic of first order, allowing strong expressions to be formulated.
Other description logics permit strong (complex) expressions in a very compact representation. For description logics, there are special inputs and intuitive notation that facilitates the handling of them substantially.
Modeling in expert systems is very important, especially in systems within highly complex domains where spatial and temporal data needs to be modeled. The ontology model permits the use of a *reasoner* that can check definitions of the statements in the ontology for consistency against each other. It can also recognize which concepts are the best for which definitions, pursuing an optimal solution in terms of size, speed, etc. This is particularly helpful when dealing with multiple classes hierarchies, therefore expert systems permit creating complex concepts from definitions of simpler ones. So, an ontology is an engineering artifact or is a formal representation of knowledge as a set of concepts within a domain, it often includes classification based information and constraints capturing the knowledge about the domain (cf. Kohler et al. (2003)).
Rule-based systems are successfully applied across a lot of domains. The interest in ontologies has become stronger to develop a common rule base that could be computed by different rule engines. This effort has led to the development of several rule languages such as the *rule markup language* (RuleML), the *semantic web rule language* (SWRL), Metalog, ISO Prolog, and many others.
Beside the weaknesses of SWRL are the weaknesses of the *SPARQL protocol and RDF Query Language* (SPARQL), where RDF is the acronym for *resource description framework* (see World Wide Web Consortium (2010)), to query an ontology, which requires the query writer to understand the data structure of the RDF resources. This understanding can be derived from *eye parsing* where sometimes the RDF or OWL ontology are large and the human being is not able to follow any more. This can become a major obstacle when debugging or extending an existing ontology.
Other computing paradigms, such as constraint satisfaction, quantified boolean formulas (QBF), or first order logic (FOL), do not naturally offer the powerful expressive possibilities to define our knowledge database of to model the spatial and context models. In general the tasks posed in the constraint satisfaction paradigm are computationally intractable (NP-hard).

Thinking of real-time reasoning systems, the most challenging task is to provide decision support in competitive situations, which calls for fast and reliable decisions. Speaking about the quality of a decision, we can hardly hope to reliably forecast another entity's behavior during an interaction. This is due to different ontologies most likely residing in different agents, as well as other unexpected behavior of a human operator who decides to ignore the system's recommendation.

1.1 Assurance

In many situations, such as driving a car with an adaptive driving assistance system (ADAS), the situation at hand is not truly competitive, but equally sure not cooperative. Even if cooperative driving is intended by the individual assistance system (presuming those to be capable of vehicle-to-vehicle ad hoc networking), the ultimate decision lies with the driver, who can freely choose to obey or to ignore any recommendation. It follows that if we wish to *assure* a particular outcome tied to a recommendation, we have to consider the behavior of others as much as our own. Interestingly, concepts from game-theory do provide an elegant mean of having a guaranteed outcome regardless of what the other entities in the system do. These concepts have become known under the general term of *security strategies* (cf. Ghose & Prasad (1989)), and can effectively be carried over to the context of ontological reasoning.

For the modeling, we shall assume an ontology with reasoning capabilities as a mere function taking some input values (constituting the query) to some recommendation (by performing internal reasoning). For illustrative purposes, let us take adaptive driving assistance as an example, stressing it whenever ideas or concept shall be exemplified. Suppose that the ADAS has been queried for driving directions, provides us with three alternative recommendations for guiding us to a given destination as fast as possible.

1. drive straight at high speed.
2. turn left at the next junction, speed can be moderate.
3. turn over with high speed.

The first recommendation may arise when the route is correct, but not necessarily the shortest one. The second advise may be appropriate if a faster route can be reached from the next junction, and the third statement could indicate an even shorter route provided that one immediately turns around.

Having the additional objective that we seek to maximize safety, one can again use the ontology and reasoning capabilities to deduce the likelihood of an accident. Assume that the approaching traffic can either come at slow or fast speed, and is free to turn left, right or go straight. We cannot hope to correctly guess which one happens, so we seek the best possible outcome in this uncertain situation.

The ontology plays a twofold role in the above scenario, as it can be used to

1. derive a recommendation for the next action, and to
2. assess the consequences that such an action might have.

While ontologies can do both, they are usually not explicitly designed to yield the "best" recommendation. Certainty factors, fuzzy reasoning and many other such approaches have been introduced to handle uncertainty, but a quantitative optimization of responses is mostly beyond the scope. Designing an ontology to provide us with both,

- correct, and
- provably optimal,

recommendations is a hardly considered problem, and a solution proposal for it is the core contribution of this chapter. In the following, we discuss both problems separately, tackling the correctness issue with a refined form of diagnosis. Optimality of recommendation will be achieved with a post-processing step, to be described in section 4.

2. Ontological reasoning in practice

The most recent developments for ontology design is *ontology web language* (OWL), with its dialects OWL DL (particularly optimized for description logics) and OWL Lite. It has various sets of operators and it is based on different logical models which simplify the description of the concepts. The semantic web rule language is based on OWL DL, OWL Lite, and the Rule Markup Language (cf. Snidaro et al. (2008)). All rules are expressed in terms of OWL concepts (classes, properties, individuals). This means that rules can be used to extract new knowledge from existing OWL ontologies. Therefore complex rules must be transformed to the requirements of SWRL (cf. Matheus et al. (2005)). Also, there are no inference engines that fully support SWRL up to now.
The first step in building a context model is to specify the desired system behavior. The developer then lists a set of possible scenarios, where each scenario is a relationship between entities to be observed. The requirements for modeling information contexts (cf. Fuchs (2008)):

Applicability:	The model must restrict the domain of application.
Traceability:	The model must provide support for recording of provenance and processing of information.
Inference:	The model should include tools that permit the definition of new contextual categories and facts on the basis of low-order context.
Re-usability:	The model should allow re-usability in other independent modeling tasks.
Flexibility:	The model should not be easily changeable to extend the ontology.
Completeness:	The model should cover all relevant concepts and properties.
Redundancy:	The model should not contain a lot of defined instances that have the same properties.

Context reasoning extends context information implicitly by introducing deduced context derived from other types of context. It is a perfect solution to resolve context inconsistency. In the light of the key role that ontologies play in many applications, it is essential to provide tools and services to support users in designing and maintaining high quality ontologies. This calls for:

1. All named classes can be instantiated (i.e. there are no "abstract" classes)
2. Correctly captured intuitions of domain experts
3. Minimal redundancy and no unintended synonyms
4. Rich axiomatization and (sufficient) detail

Answer queries over ontology classes and instances, e.g.:

1. Find more general/specific classes
2. Retrieve individuals/tuples matching a given query

Context interpreters consist of context reasoning engines and context knowledge-bases (context KB). The context reasoning engines provide the inference services including inferring contexts, resolving context conflicts (basically a problem of *diagnostics*) and maintaining the consistency of context knowledge-bases. Different rules for consistency can be specified and

fed into the reasoning engines. The context KB provides the service that other components can query, add, delete or modify context knowledge stored in the context database (cf. Wang et al. (2004)).
However, the problem of the current rules and reasoning systems is that they do not offer high performance according the highly declarative way of the encoding of the problem and the execution time to find the "best" solution.
As an example the used semantic web rule language (SWRL) has no negation or disjunction and the arithmetic predicates which it offers are weak.

Obstacles in writing logical programs

Semantic web technology is widely used for reasoning and expert systems. Especially the use of SWRL to define relationships between classes and individuals in the ontology may cause major problems due to the long execution time of semantic web technology for querying the ontology (e.g. via SWRL and SPARQL). This is indeed a problem when creating practical expert systems ought to perform in real-time. Some existing paradigms as well suffer from negation as failure, which has been discussed by Matheus et al. (2005). Moreover, they not necessarily allow n-ary predicates within a rule.
Going away from SWRL, we still require constraints, negation as failure, and expressive representation that are decidable and permit reasoning (efficiently). In particular, some logical programming languages do not offer arithmetic operations "built-in", and numeric constraints can affect decidability. Since reasoning (like done in the Prolog-language for instance), is often a recursive procedure, its complexity can become exponential and thus infeasible for even a medium-size ontology.
Now, to avoid the weaknesses of the existing paradigms *answer set programming* (ASP) offers flexible and high performance reasoning. ASP is a declarative approach for modeling and solves search problems by representing them as logic programs.

2.1 Answer set programming

The importance of ASP lies in the fact that it provides meaning to logic programs with default negation "not". Many interesting applications exist in planning, reasoning about action, configuration, diagnosis, space shuttle control, spatial, temporal and probabilistic reasoning, constraint programming, etc.
The Technical University of Vienna (TU-Wien) hosts a the research group "knowledge based systems", whose members are running a project on "Answer Set Programming for the Semantic Web". The goal of this project is research towards methods for providing advanced reasoning services in the context of the Semantic Web, using declarative knowledge representation and reasoning techniques (see Eiter et al. (2005)).
A logic program in the language of *AnsProlog* (also known as *A-Prolog*) is a set of rules of the form:

$$a_0 \leftarrow a_1, \ldots, a_m, \text{not}\, a_{m+1}, \ldots, \text{not}\, a_n \tag{1}$$

where $0 \leqslant m \leqslant n$, each a_i is an atom of some propositional language and *not* represents *negation-as-failure*. A negation-as-failure literal (or *naf-literal*) has the form not a, where a is an atom. Given a rule of this form, the left and right hand sides are called the *head* and *body*, respectively. A rule may have either an empty head or an empty body, but not both. Rules with an empty head are called constraints, while those with an empty body are known as *facts*. A *definite rule* is a rule which does not contain naf-literals, and a *definite program* is solely composed of definite rules (cf. Baral et al. (2010)).

Let X be a set of ground atoms (i.e. all atoms constructed with the predicate in Herband base of a logic program). The body of a rule of the form (1) is satisfied by X if $\{a_{m+1}, \ldots, a_n\} \bigcap X = \emptyset$ and $\{a_1, \ldots, a_m\} \subseteq X$. A rule with a non-empty head is satisfied by X if either its body is not satisfied by X, or $a_0 \in X$. A constraint is satisfied by X if its body is not satisfied by X.
Since logic programs unify declarative and procedural representations of knowledge, one way to reason is by using Horn clauses, backward reasoning and selective linear definite clause (SLD) resolution. The *reduct* of a program is a possibility to generate answer sets. Given an arbitrary program, Π and a set of ground atoms, X, the reduct of Π w.r.t. X, Π^X, is the definite program obtained from the set of all ground instances of Π by:

1. deleting all the rules that have a naf-literal not a in the body where $a \in X$, and
2. removing all naf-literals in the bodies of the remaining rules.

A set of ground atoms X is an answer set of a program Π, if it satisfies the following conditions:

1. If Π is a definite program, then X is a minimal set of atoms that satisfies all the rules in Π.
2. If Π is not a definite program, then X is the answer set of Π^X. (Recall that Π^X is a definite program, and its answer set is defined in the first item (cf. Baral et al. (2010)).

The other advantage of ASP is that the order of program rules does not matter and the order of subgoals in a rule is also not relevant. For an example, if we have the famous problem "3-colorability", where we have a map and we want to check whether 3 colors (blue, yellow and red) are sufficient to color a map. A map is represented by a graph, with facts about nodes and edges.

```
vertex(a), vertex(b), edge(a,b).
```

Every vertex must be colored with exactly one color:

```
color(V,r) :- vertex(V), not color(V,b), not color(V,y).
color(V,b) :- vertex(V), not color(V,r), not color(V,y).
color(V,y) :- vertex(V), not color(V,b), not color(V,r).
```

No adjacent vertices may be colored with the same color

```
:- vertex(V), vertex(U), edge(V,U), col(C), color(V,C), color(U,C).
```

Of course, we need to say what colors are:

```
col(r).
col(b).
col(y).
```

After running this program we will get all possible coloring cases to color the whole map with three different colors. The other advantage of ASP that the order of program rules does not a matter and the order of subgoals in a rule does not a matter also.

2.1.1 Logic programming with ordered disjunction

Logic programming can be extended to allow us to represent new options for problems in the head of the rules. ASP gives us this ability by the way of ordered disjunctions. Using ASP under specific conditions reasoning from most preferred answer sets gives optimal problem solutions. Through logical programs with ordered disjunction (LPODs), such as normal logic programs we are able to express incomplete and defeasible knowledge through the use of default negation, they allow us to represent performances among intended properties of problem solutions which depend on the current context. It is possible to use the degree of satisfaction of a rule to define a preference relation on answer sets. We will present an alternative on game-theoretic grounds in section 4. Brewka (2002) defines a rule as having

degree 1 under the following condition: when A is an answer set of P, then A satisfies all rules of P. For example, let us plan a vacation: Normally you prefer to go to Mallorca but also to Stockholm (denoted by the preference relation $\prec$). Usually people prefer Mallorca over Stockholm, unless it is hot. If it is hot Mallorca is preferred over Stockholm. In summer it is normally hot, but there are exceptions. If it is winter, then Mallorca is no long considered (cf. Brewka (2002)).

$$\begin{aligned} Stockholm \prec Mallorca &\leftarrow \text{not}\, hot && \text{(rule 1)} \\ Mallorca \prec Stockholm &\leftarrow hot && \text{(rule 2)} \\ hot &\leftarrow \text{not}\, \neg hot, summer && \text{(rule 3)} \\ \neg Mallorca &\leftarrow rain && \text{(rule 4)} \end{aligned}$$

Without further information about the weather we obtain the single preferred answer set $A_1 = \{Stockholm\}$, there is no information that it might be *hot*, so rule 1 will determine preferences. A_1 satisfies all rules to degree 1. Now if we add a new fact *summer*, then the new answer set is $\{summer, hot, Mallorca\}$. If we add the literal *hot*, then the new answer set is $\{summer, \neg hot, Stockholm\}$. Finally, if we add the facts *summer* and *rain*. The single answer set is $\{summer, rain, hot, \neg Mallorca, Stockholm\}$, we see that it is not possible to satisfy all rules to degree 1. As in real life there are situations where the best options simply do not work out, there for LPODs are very well suited for representing problems where a certain choice has to be made. In general, using ASP we can optimize the solution we want to generate, we can improve the rules and define the constraints we are using to get the maximum optimization of the desired answer sets (solutions) (cf. Brewka (2002)). Assurance, as introduced in section 4, pursues similar goals.

2.1.2 Guess and check programs in answer set programming

Answer set programming (ASP) is widely used, expressing properties in NP (i.e. properties whose verification can be done in polynomial time), where answer sets of normal logic programs can be generated through solutions and polynomial time proofs for such properties. The solution of such problems can be carried out in two steps:

1. Generate a candidate solution through a logic program
2. Check the solution by another logic program (cf. Eiter & Polleres (2006))

However, it is often not clear how to combine Π_{guess} and Π_{check} into a single program Π_{solve} which solves the overall problem. If we simply take the union $\Pi_{guess} \vee \Pi_{solve}$ does not work, so we have to rewrite the program.
Theoretical results prove that for problems with Σ_2^P complexity, it is required that Π_{check} is rewritten into a disjunctive logic program $\grave{\Pi}_{check}$ such that the answer sets of $\Pi_{solve} = \Pi_{guess} \vee \grave{\Pi}_{check}$ yield the solutions of the problem, where $\grave{\Pi}_{check}$ emulates the inconsistency check for $\grave{\Pi}_{check}$ as a minimal model check, which is co-NP-complete for disjunctive programs. This becomes even more complicated by the fact that $\grave{\Pi}_{check}$ must not crucially rely the use of negation, since it is essentially determined by the Π_{guess} part. These difficulties can make rewriting Π_{check} to $\grave{\Pi}_{check}$ a formidable and challenging task (cf. Eiter & Polleres (2006)).
As an example, if we are talking about planning the problem to find a sequence of actions, which it takes the system from an initial state p_0 to a state p_n, where the states are changing over the time. Conformant planning looks for a plan L which works under all contingencies cases that may be caused by incomplete information about the initial state and/or nondeterministic actions effects which is Σ_2^P under certain restrictions (see Eiter & Polleres (2006)). We consider the problem of the"fire alarm", we have an alarm that there is a

fire in a building which is supported through a fire alarm system. Possible actions (states) of the system turn off the electricity and then to pump water. After just having turned off the electricity, it does not extinguish the fire, but only the pumping of water guarantees that it is really extinguished. Using the following guess and check programs $fire_{guess}$ and $fire_{check}$ respectively, we can compute a plan for extinguishing the fire by two actions, $fire_{guess}$ and $fire_{check}$, the program $fire_{guess}$ guesses all candidate plans $P = p_1, p_2, ..., p_n$ using time points for action execution,

```
fire_guess:
% Timestamps:
time(0).
time(1).
% Guess a plan:
turn_off(T) v -pump(T) :- time(T).
pump(T) v -pump(T) :- time(T).
% Forbid concurrent actions:
:- pump(T), turn_off(T).
```

while $fire_{check}$ checks whether any such plan P is conformant for the goal g = not $extinguished(2)$ The final constraint eliminates a plan execution if it reaches the goal; thus, $fire_{check}$ has no answer set if the plan P is conformant.

```
fire_check:
% Initial state:
fired(0) v -fired(0).
% Frame Axioms:
fired(T1) :- fired(T),
             time(T),
             not -fired(T1),
             T1 = T + 1.
turned_off(T1) :- turn_off(T),
                  T1 = T + 1.
% Effect of turning off:
turned_off(T1) :- turn_off(T),
                  T1 = T + 1.
fired(T1) v -fired(T1) :- turn_off(T),
                          fired(T),
                          T1 = T + 1.
% Effect of pumping:
-fired(T1) :- pump(T),
              turn_off(T),
              T1 = T + 1.
% Check goal in stage 2 (constraint):
:- not fired(2).
```

The program $fire_{guess}$ generates the answer set $S = \{time(0), time(1), turn_{off}(0), pump(1)\}$ which corresponds to the (single) conformant plan $\{P = turn_{off}, pump\}$ for goal not $fired(2)$. Using the method $fire_{guess}$ and $fire_{check}$ can be integrated automatically into a single program $fire_{solve} = fire_{guess} \vee fire_{check}$ It has a single answer set, corresponding to the single conformant plan $P = \{turn_{off}, pump\}$ as desired.

With these examples in mind, we now turn to the problem of diagnosing such ontologies. As should have become evident by now, spotting an error in a large-scale program is a challenging task. We deliver a solution that is flexible and can be implemented with widely standard components. In particular, our proposal does not require substantial changes to an

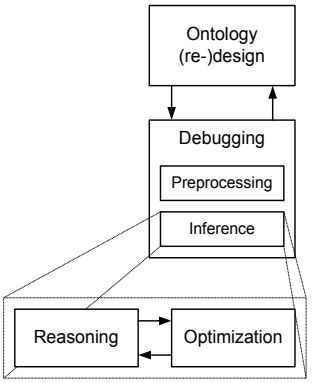


Fig. 1. Ontology design

existing diagnostic engine, so it can be seen as an "add-on" or refinement of a debugging system.

3. Fine-grained axiom diagnosis

In order to generalize diagnostic methods to fine-grained axiom diagnosis, we review the concepts behind model-based diagnosis. This prepares the ground for the diagnostic algorithm that concludes this section. The workflow when designing and debugging an ontology is briefly sketched in figure 1. Our focus in this section will be on debugging.
We assume the reader familiar with the resolution principle and predicate and propositional logic. For convenience, however, we will briefly review some basics and results before going into technical details (see (de Wolf & Nienhuys-Cheng, 1997, Chapter 2) for full details).
Let ψ be a first-order logic formula. An *interpretation* of ψ is an assignment of the variables to values from a given domain, and an assignment of predicates over this domain to truth-values, such that ψ becomes true. We write lit(Σ) for the set of literals appearing in either a formula Σ or a set Σ of formulas. For a set of formulas Σ, a *model* is an interpretation such that every $\psi \in \Sigma$ is true. Let Σ be a set of formulas and let ψ be a formula. We say that Σ logically implies ψ, writing $\Sigma \models \psi$, if every model of Σ is also a model of ψ. For sets of formulas, we write $\Sigma \models \Gamma$, if $\Sigma \models \psi$ for every $\psi \in \Gamma$.
A *clause* is basically a logical disjunction of literals. Let two clauses $C_1 = L_1 \vee \ldots \vee L_m$ and $C_2 = M_1 \vee \ldots \vee M_n$ be given and assume that there are two terms L_i and M_j that can be unified with each other such that $L_i = \neg M_j$. The *resolvent* of C_1, C_2 is the expression $C_1 \vee C_2$, having the terms L_i and M_j omitted. For the upcoming results, it is not necessary to fully introduce the concept of clauses and resolution, and we will confine ourselves to the following informal example of resolution: assume that we know that Peter plays either chess or football (clause C_1). In addition, we know that he does not play chess (clause C_2). We conclude that he must play football (the resolvent is thus $C_1 \vee C_2$ without the assertion of Peter playing chess).
For a clause C, we say that it can be *derived* from a set Σ, if there is a finite sequence of clauses $R_1, \ldots, R_n = C$ such that each clause R_i is either in Σ or a resolvent of two (previous) clauses R_j, R_k with $j, k < i$. In that case, we write $\Sigma \vdash_r C$.
Finally, we denote the empty clause as $\perp$, and note that a set of sentences is *inconsistent*, if it models the empty clause. Hence, writing $\Sigma \models \perp$ is equivalent to saying that a contradiction is derivable from Σ.

3.1 The general diagnosis problem

The theory of diagnosis as employed in this work is based on the seminal paper of Reiter (1987), with corrections made by Greiner et al. (1989). The method devised in the sequel proves axiom pinpointing to be a mere special case of standard diagnosis, except for some preprocessing. Hence, competing alternative approaches to the same problem, such as contrived by Schlobach & Cornet (2003) for instance, are technically interesting but come at the disadvantage of calling for intricate extensions to a diagnostic engine. Our approach comes at negligible additional computational cost, as we will need a parser for the axiom syntax (which cannot be avoided anyway).

To reason within an ontology is the process of deriving assertions about individuals from known facts and rules. Occasionally, we have certain facts and rules in ϕ that are undoubtedly correct and consistent, and we collect these in a set $B \subseteq \phi$, calling it the *consistent background theory*. These are explicitly excluded from the diagnostic process. The remainder $KB = \phi \setminus B$ is the set of facts and rules that are subject to the diagnosis. An ontology is usually verified by testing it against positive and negative test-cases.

Informally, a *diagnosis* D is a minimal set of components whose replacement will create a consistent ontology. This has been formally captured by Friedrich & Shchekotykhin (2005), and is clearly defined in

Definition 3.1 (Diagnosis (Problem))**.** *A* diagnosis problem *is a quadruple* (KB, B, TC^+, TC^-)*, where KB is an ontology, comprising a set of facts and rules to be diagnosed, B is a consistent set of facts and rules (called the* background theory*, and* TC^+ *and* TC^- *are positive and negative test-cases. A* diagnosis $D \subset KB$ *is a minimal set (in terms of the subset-relation) of sentences such that there is an extension EX, where EX is a set of logical sentences added to KB, such that*

1. $\forall e^+ \in TC^+ : (KB \setminus D) \cup B \cup EX \cup \{e^+\} \not\models \bot$,
2. $\forall e^- \in TC^- : (KB \setminus D) \cup B \cup EX \cup \{e^-\} \models \bot$.

Rephrasing this definition, in essence it states that by replacing the components in the diagnosis D with the components in EX makes the ontology ϕ consistent with every positive test-case and inconsistent with every negative test-case.

Example: for illustration, consider the following example ontology:

$\mathcal{A}_1 : A_1 \rightarrow \neg A \wedge A_2 \wedge A_3$	$\mathcal{A}_2 : A_2 \rightarrow \neg D \wedge A_4$
$\mathcal{A}_3 : A_3 \rightarrow A_4 \wedge A_5$	$\mathcal{A}_4 : A_4 \rightarrow \forall s : F \wedge C$
$\mathcal{A}_5 : A_5 \rightarrow \exists s : \neg F$	$\mathcal{A}_6 : A_6 \rightarrow A_4 \wedge D$

The inconsistency is revealed by the following reasoning, starting with A_1 as a known fact: from $\mathcal{A}_1$, we can deduce (among others) the assertion that A_3 holds true. This one in turn implies the validity of A_4 and A_5 via rule $\mathcal{A}_3$. Yet A_4 tells that all individuals s enjoy property F, while A_5 implies that there is at least one individual that does *not* fall into the class F. This inconsistency is to be removed with the aid of diagnosis.

The problem tackled in the following is motivated by the (initially made) observation that the process will not remain that simple in a real-life ontology. Facing very complicated axioms, the diagnostic algorithms inspired by the work of Reiter (1987), essentially provide rather coarse-grained pointers towards the error. We shall improve on this.

Computing a diagnosis relies on the concept of *conflict sets* as introduced by de Kleer (1976) and adapted by Reiter (1987). Given a diagnosis problem (KB, B, TC^+, TC^-), a *conflict set* $C \subseteq KB$ is such that $C \cup B$ is inconsistent (either by itself or with any of the test-cases). A conflict is said to be minimal, if no subset of C is a conflict. Computing conflict sets can be done

with a divide-and-conquer approach, and we refer the reader to the QUICKXPLAIN-algorithm by Junker (2004) for that matter.
Returning to the example, for illustrative purposes, one conflict for the ontology is found as:

$$C = \{\mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}\,.$$

Removing the set C from the ontology makes everything consistent again. It is easy to construct examples where other conflicts can be found as well.
Computing a diagnosis then amounts to computing a hitting-set for the collection of all conflicts that exist within KB. This is basically theorem 4.4 in Reiter (1987). This is not surprising and easily justified intuitively: assume D to be a hitting-set for the set of minimal conflicts. Then retracting D from KB will reduce every conflict set by at least one element, thus destroying the conflict because those sets are minimal. Since this approach to diagnosis has seen widespread implementation and many appropriate systems are in place, we leave the details of the computation aside and concentrate on the problem of axiom pinpointing, which arises from a shortcoming of the so-far described method.

3.2 Pinpointing erroneous parts of an axiom

Despite the elegance of the existing diagnosis theory, its accuracy is limited to pointing towards entire axioms. Hence these can be rather complicated and lengthy, which makes the actual debugging process, i.e. finding the replacement EX (cf. definition 3.1) actually hard. Consequently, we would like to extend the scope of diagnosis to the smallest building blocks of the sentences within ϕ. This problem is known as *axiom diagnosis*. The trick is using the syntactical structure of a sentence to decompose a complicated axiom into a sequence of simple axioms representing only single logical connectives, quantifiers, etc. We will call this an *irreducible decomposition*, and the whole process of axiom pinpointing then boils down to diagnosing such an irreducible decomposition of the axiom to be debugged. We will expand the details in this section, strongly drawing from the work of Friedrich et al. (2006), whilst presenting further theoretical results that do not appear in this reference.
To ease technicalities in the following, we need two technical yet mild assumptions:

- axioms obey a non-ambiguous formal grammar G, such that every application of a production rule produces only *one* new logical connective or quantifier in each step of the derivation. We denote the language induced by the grammar G as $L(G)$, and assume $\mathcal{A} \in L(G)$ for every axiom $\mathcal{A}$, i.e. it is generated by G in a non-ambiguous manner. Although the ambiguity decision problem is undecidable for general grammars, those permitting disambiguation by lookahead tokens are provably unambiguous and thus suitable for our needs.
- the *open world assumption* states that a proposition is true if it can be proven, and assumed to be false if it can be disproved. Any non-present proposition remains unknown until proven to be true or false. In other words, reasoning under the open-world assumption demands first making choices before deducing anything.

Our first result formally captures the observation that the parse-tree of an axiom naturally yields a decomposition that only consists of axioms of the simplest possible form.
Let $\mathcal{A}$ be an axiom with its associated parse-tree. As an example, the axiom $A \to (B \wedge C) \vee D$ can be decomposed into

$$\zeta(\mathcal{A}) = \{A \to X_1 \vee X_2, X_1 \to X_3 \wedge X_4, X_3 \equiv B, X_4 \equiv C, X_2 \equiv D\}\,.$$

In this sense, an "atomic" axiom in the above sense, i.e. an axiom $\mathcal{A}$ for which the set $\zeta(\mathcal{A})$ is singleton, is said to be *irreducible*. For example, the formula $A \wedge B$ is irreducible, as it involves only a single connective between two literals. On the contrary, the formula $A \to \exists x : p(x) \wedge q(x)$ involves a quantifier, an implication and two unary predicates, and is therefore reducible. This formula can be decomposed into

$$\zeta(\mathcal{A}) = \{A \to X, X \to \exists x : R(x), R(x) \equiv P(x) \wedge Q(x), P \equiv p(x), Q \equiv q(x)\},$$

which is a set of irreducible axioms. For sets C of axioms, we define $\zeta(C) = \bigcup_{\mathcal{A} \in C} \zeta(\mathcal{A})$. Such a decomposition is called *irreducible*, if each of its elements is irreducible.
For example, the statement

```
p(A) :- (q(A,B), r(B)) ; s(A)
```

is irreducibly decomposed into the set ζ being

$$\begin{array}{llllll} \{X_0 \to X_1, & X_0 \equiv p(X_2), & X_2 \equiv A, & X_1 \equiv X_3 \vee X_4, & X_3 \equiv X_5 \wedge X_6, & X_5 \equiv q(X_7, X_8), \\ X_7 \equiv A, & X_8 \equiv B, & X_6 \equiv r(X_9), & X_9 \equiv B, & X_4 \equiv s(X_{10}), & X_{10} \equiv A\}. \end{array} \tag{2}$$

In the light of these considerations, the next result is immediately clear:

Lemma 3.1. *Let G be a non-ambiguous grammar satisfying the assumption above, and let $\mathcal{A}$ denote an axiom in $L(G)$. Then there is a unique set of axioms of the form $p(X, Y_1, \ldots, Y_n)$, where p is a logical connective or a quantifier, and $X, Y_1, \ldots, Y_n$ are literals.*

We denote the decomposition implied by the lemma 3.1 by $\zeta(\mathcal{A})$. It is quite obvious that $\zeta(\mathcal{A})$ is logically equivalent to $\mathcal{A}$, as we can recover $\mathcal{A}$ by back-substituting the literals, which is equivalent to stating that $\zeta(\mathcal{A}) \models \mathcal{A}$ and vice versa.
To ease notation, for logical implications or equivalences like $\mathcal{A} : X \to p(Y_1, \ldots, Y_n)$ or $\mathcal{A} : X \equiv p(Y_1, \ldots, Y_n)$, we write $lhs(\mathcal{A})$ to denote the set $\{X\}$, and $rhs(\mathcal{A})$ to collect all literals appearing on the right-hand side of $\mathcal{A}$, i.e. $rhs(\mathcal{A}) = \{Y_1, \ldots, Y_n\}$. Returning to the example above, we would have $lhs(X_0 \to X_1) = \{X_0\}$ and $rhs(X_0 \to X_1) = \{X_1\}$. For a set C of axioms, we set $lhs(C) = \bigcup_{\mathcal{A} \in C} lhs(\mathcal{A})$. The symbol $rhs(C)$ is defined analogously.
As a mere technical tool, we introduce a graphic representation of the structure of an axiom. It is closely related to the concept of parse-trees:

Definition 3.2 (structure graph)**.** *Let $\zeta(\mathcal{A})$ be an irreducible decomposition of an axiom. The vertices of the* structure graph *$G^s_{\mathcal{A}}$ are given by $V(G^s_{\mathcal{A}}) = lhs(\zeta(\mathcal{A}))$, i.e. the vertices are the names of all axioms that appear in $\zeta(\mathcal{A})$. The graph has an edge between two axioms $\mathcal{A}_i$ and $\mathcal{A}_j$, if and only if $rhs(\mathcal{A}_i) \cap lhs(\mathcal{A}_j) \neq \emptyset$.*

Graphically, the decomposition (2) would look like shown in figure 2. The vertices of the structure graph are given by $lhs(\zeta) = \{X_0, \ldots, X_{10}\}$, reading off the left-hand sides from the decomposition (2). An arc is present between two vertices, if they appear on different sides of the corresponding axiom, i.e. X_0 and X_1 are connected because the decomposition contains the rule "$X_0 \to X_1$".
This provides us with an adequate tool for simplifying axioms before putting it to the diagnosis. Indeed, the structure graph's connectivity can be related to logical inconsistency:

Lemma 3.2 ((Rass, 2005, Proposition 5.3.3))**.** *Let C be a minimal conflict of an irreducible axiom decomposition $\zeta(C)$. Then the corresponding structure graph is connected.*

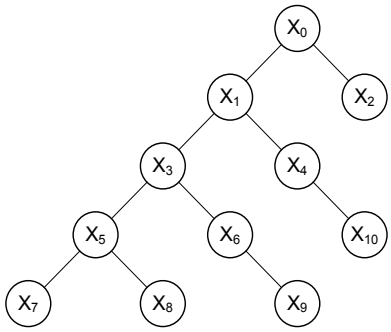


Fig. 2. Structure graph for decomposition ζ (eq. (2))

Proof. By construction, an irreducible axiom decomposition is tree-structured, so let X_0 denote the top-level literal. Assume the proposition to be false: Let C be a minimal conflict and let the corresponding structure graph be disconnected, having two non-empty connectivity components C_1 and C_2 ($C = C_1 \cup C_2$). Let S, S_1 and S_2 be standard forms of $\mathcal{B}, C_1$ and C_2, respectively. Then, by (de Wolf & Nienhuys-Cheng, 1997, theorem 5.20), we have

$$\mathcal{B} \cup C \models \bot \iff S \cup S_1 \cup S_2 \vdash_r \bot.$$

By definition, this is the case if and only if

$$S \cup S_1 \cup S_2 \vdash_r \{R_1, R_2\} \vdash_r \bot.$$

Since C is a minimal conflict, $\mathcal{B} \cup C_1$ is consistent as well as $\mathcal{B} \cup C_2$. Hence, we cannot have both R_1 and R_2 derivable from the same connectivity component, say C_1, for otherwise

$$S \cup S_1 \vdash_r \{R_1, R_2\} \vdash_r \bot$$

and therefore $\mathcal{B} \cup C_1 \models \bot$ by (de Wolf & Nienhuys-Cheng, 1997, theorem 4.39). This would mean that C_1 is a smaller conflict set than C, contradicting the minimality of C. So assume without loss of generality that

$$S \cup S_1 \vdash_r R_1 \text{ and } S \cup S_2 \vdash_r R_2. \tag{3}$$

By construction, there must be a sequence $R'_1, R'_2, \ldots, R'_n \equiv R_1$ with $R'_i \in S \cup S_1$ or R'_i a resolvent of $\left\{R'_p, R'_q\right\}$ for $p, q < i$. Since $\mathcal{B}$ is consistent, $S \nvdash_r R_1$ and there must be an R'_j so that a (leaf-)literal L_1 in S_1 occurs in R'_j, as no set of interior literals can solely contribute to the derivation of R_1. Hence we must have a path from X_0 to $X_k \equiv L_1$ by our assumptions on the grammar and the method of construction, for otherwise X_k (and therefore L_1) would not occur anywhere during the inference. The same argument applies for the derivation of R_2 from $\mathcal{B} \cup C_2$, so any inference giving $\{R_1, R_2\}$ necessarily induces paths from X_0 to the node $X_k \equiv L_1 \in \text{lit}(R'_1, R'_2, \ldots, R_1)$ and from X_0 to $X_l \equiv L_2 \in \text{lit}(R''_1, R''_2, \ldots, R_2)$. Yet $X_0 \in C_1$ (without loss of generality), and C_1 is not connected to C_2 in the structure graph, so the path from X_0 to X_l does not exist. This contradicts (3) and completes the proof. □

Based on the previous result, we can state an interesting criterion that permits further minimizing conflict sets in terms of literals, even when they are already minimal in terms of set inclusion.

Proposition 3.3 ((Rass, 2005, Lemma 5.3.4)). *Let $\mathcal{K}$ be a knowledge-base and C be a minimal and irreducible conflict-set, and let $\mathcal{A} \in C$ be an axiom involving the literals $X, Y_1, \ldots, Y_n$. Then X can be removed without loosing the minimality or inconsistency if, and only if,*

$$X \in (rhs(C) \setminus lhs(C)) \cap \overline{lit(\mathcal{K})}.$$

Proof. ("if") Let $C = \{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ be inconsistent, that is $B \cup C \models \bot$ for some consistent background theory B. Without loss of generality, we assume $\mathcal{A}_1 : p(\ldots, X, \ldots)$ to be the axiom that shall be simplified by removing X. Then it is clear that

$$B \cup (C \setminus \{\mathcal{A}_1\}) \cup \{\mathcal{A}_1\} \models \bot. \tag{4}$$

Since the literal X neither appears on the left-hand side of any axiom nor anywhere else in the knowledge-base, we do not have any specification or information on it. Thus, by the open-world assumption, equation (4) does especially hold if we assume an arbitrary but fixed value c for X. Let c be the neutral element for the operator p (which would be `false` if $p = \vee$ or `true` if $p = \wedge$, respectively). By setting $X \equiv c$ we get

$$B \cup [(C \setminus \{\mathcal{A}_1\}) \cup \{\mathcal{A}_1\}] \cup \{X \equiv c\} \models \bot,$$

which can be rewritten as

$$B \cup (C \setminus \{\mathcal{A}_1\}) \cup (\{\mathcal{A}_1\} \cup \{X \equiv c\}) \models \bot,$$

being equivalent to

$$B \cup (C \setminus \{\mathcal{A}_1\}) \cup \{\mathcal{A}_1^*\} \models \bot$$

with $\mathcal{A}_1^*$ being the axiom $\mathcal{A}_1$ having X removed.
If C is minimal, then so is $(C \setminus \{\mathcal{A}_1\}) \cup \{\mathcal{A}_1^*\}$, because removing $\mathcal{A}_1^*$ results in the set $C \setminus \{\mathcal{A}_1\}$, which is consistent due to the minimality of C.
("only if") we show that removing a literal in

$$\overline{\left((rhs(C) \setminus lhs(C)) \cap \overline{\mathrm{lit}(\mathcal{K})}\right)} = \overline{rhs(C)} \cup lhs(C) \cup \mathrm{lit}(\mathcal{K})$$

destroys either the conflict or the minimality: by the construction of an irreducible decomposition as used in lemma 3.1, the structure graph induced by C is a tree. Assume the existence of another axiom $\mathcal{A}_k : X \equiv q(\ldots) \in C$ and that the set is still a minimal conflict after having removed X from the right hand side of $\mathcal{A}_1$. However, if we do so, then the edge between $\mathcal{A}_1$ and $\mathcal{A}_k$ disappears and the tree breaks up into two non-connected components. By lemma 3.2, the resulting set is not a minimal conflict any more, thus contradicting our assumption. Furthermore, we cannot remove any literal $L \in \mathrm{lit}(\mathcal{K})$, because the construction employed by lemma 3.1 implies that all axioms having L appear on their right hand side are of the form $Y \equiv L$. This does not involve any operator so no further simplification is possible. □

Example: for illustrating the last result, take the set of three (inconsistent) rules

$$C = \{X_1 \to X_2 \wedge X_3 \wedge X_4, X_2 \to B, X_3 \to \neg B\}.$$

They form a minimal conflict set for some axiom decomposition. The literal X_4 appears on the right-hand sides of C, but not on the left-hand side, since we have $lhs(C) = \{X_1, X_2, X_3\}$

and $rhs(C) = \{X_2, X_3, X_4, B\}$. Since the literals X_i were auxiliary and introduced merely for the decomposition, the literal X_4 can be unhesitatingly deleted. This results in a simplified set

$$C_{\text{simplified}} = \{X_1 \to X_2 \wedge X_3, X_2 \to B, X_3 \to \neg B\},$$

which is still inconsistent for the same reasons as C is.

3.3 The axiom diagnosis algorithm

With all these preliminaries, the axiom pinpointing algorithm is essentially a standard diagnosis with preprocessing. It comprises two stages:

1. Decompose the given set of axioms $\mathcal{A}_1, \ldots, \mathcal{A}_n$ irreducibly into the set $KB = \bigcup_{i=1}^{n} \zeta(\mathcal{A}_i)$.
2. Run a standard diagnosis on KB.

It is known that diagnosis is a computationally hard problem (due to the need of reasoning for computing the conflict sets and the computation of hitting sets, which is an NP-complete problem in general Garey & Johnson (1979)). While the algorithm above has been stated previously in Friedrich et al. (2006), this preliminary work does not exploit the potential simplification provided by proposition 3.3, by instantly drawing away the attention from literals that cannot contribute to the inconsistency.
The idea of decomposing an axiom into (irreducible) sub-axioms permits controlling the granularity as well as the scope of the diagnosis most easily.

Controlling the granularity: We are free to define the production rules of the underlying grammar in such a way that more complicated axioms can be generated in one blow. A simpler, yet equally effective, approach is backsubstituting a selection of irreducible axioms, thus creating more complicated expressions, yet remaining structurally simpler than the original axiom. Returning to the previous example, one could backsubstitute some of the axioms in the decomposition (2) in order to obtain the (logically equivalent but less fine-grained) decomposition

$$\begin{aligned}&\{X_0 \to X_1, \quad X_0 \equiv p(X_2), \quad X_2 \equiv A, \quad X_1 \equiv X_3 \vee X_4,\\ &X_3 \equiv X_5 \wedge X_6,\ X_5 \equiv q(A,B),\ X_6 \equiv r(B),\ X_4 \equiv s(A)\}.\end{aligned}$$

Diagnosing the latter decomposition obviously provides us with less precise information than diagnosing the full decomposition (2).

Controlling the scope: If some parts of the axiom are correct beyond doubt, then we can simply shift these to the background theory $\mathcal{B}$ when running the diagnostic engine. Hence, the scope comprises all axioms in KB, explicitly excluding those in $\mathcal{B}$.

Example: Reiter's diagnostic algorithm, when applied to the example ontology given in section 3.1 returns the following three diagnoses:

1. $D_1 = \{A_3 \to A_4 \wedge A_5\}$
2. $D_2 = \{A_4 \to \forall s : F \wedge C\}$
3. $D_3 = \{A_5 \to \exists s : \neg F\}$

which just says that we may change one of three axioms in order to achieve a repair. Taking the first of these diagnoses D_1 and performing axiom-diagnosis on a decomposition of the axiom provides us with the following diagnoses (among others):

- $X_4 \equiv A_5$: This directly points at something wrong with the literal A_5 in the axiom. Indeed, A_5 permits deriving the expression $\exists s : \neg F$ which yields an inconsistency in connection with $\forall s : F$ (derived via A_4).
- $X_2 \equiv X_3 \wedge X_4$: This points at the $\wedge$-operator as possible flaw. This is correct, since if we replace the $\wedge$ by a $\vee$, we can easily avoid the inconsistency. Moreover, the diagnosis showed that both arguments of the $\wedge$ are necessary for the inconsistency, which is also correct as we have to be able to conclude $\mathcal{A}_4$ and $\mathcal{A}_5$ from $\mathcal{A}_3$ to be inconsistent.

Notice the appeal of this form of refined diagnosis, as the precise location of the error is actually marked with an "X". By then, it is up to the human engineer to do the repair properly.

4. Assurance

Suppose that an ontology upon receiving the query has presented us with a number of, say n_1, answers from which we can choose one. From the viewpoint of the ontology, each answer is logically correct, and in the absence of preferences, certainty factors, or other means of selection, we can only choose the best one subjectively. The goal of assurance is making this decision with a provable benefit. For that matter, we briefly introduce some elements from the theory of games, which will become handy when putting things together to a *reasoning engine with assurance*. The reader familiar with matrix games may safely skip section 4.1, and move on to section 4.2 directly.

4.1 Matrix-games

A (*non-cooperative n-person*) *game* $\Gamma = (N, PS, H)$ is a triple composed of a set $N = \{1, 2, \ldots, n\}$ of players being able to choose actions from their corresponding strategies within the set of sets $PS = \{PS_1, PS_2, \ldots, PS_n\}$. The i-th player, by taking action $s_i \in PS_i$ from his set PS_i of possible strategies, receives the payoff $u_i(s_i, s_{-i})$, where $u_i \in H$ and s_{-i} denotes the strategies chosen by i's opponents. The set H thus comprises the set of payoff functions for each player, i.e. $H = \{u_i | u_i : \times_{i=1}^{n} PS_i \to \mathbb{R}\}$. Although we will use the general definition here, our application use of game-theory will be with 2-player games, with player 1 being the user of the ontology, and player 2 being the whole set of remaining entities outside the user's scope.
A *(Nash-)equilibrium* is a choice $s^* = (s_1^*, \ldots, s_n^*)$ such that

$$u(s_i, s_{-i}^*) \leq u(s_i^*, s_{-i}^*) \quad \forall s_i \in PS_i$$

and for all $i \in N$, i.e. if any of the players solely chooses an action other than s_i^*, his revenue will decrease. It is easy to construct examples where equilibria are not existing among the pure strategies in PS_i. But, if strategies are understood as probabilities for taking certain actions during repetitions of the game, then Glicksberg (1952) has proven that equilibria exist for every game with continuous payoff functions. In that case, the payoff is averaged over the repetitions of the game, i.e. we consider the *expected payoff*. Strategies which are interpreted as probability distributions over the sets in PS are called *mixed strategies*, and we shall exclusively refer to these in the sequel. The set S_i consists of all mixed strategies over PS_i. A game is called *zero-sum*, if $\sum_i u_i = 0$, or in the two-person case, if $u_1 = -u_2$. The game is called *finite*, if the sets in PS are all finite. For a finite zero-sum game Γ_0, the average revenue under an equilibrium strategy is the *value* of the game, denoted as $v(\Gamma_0) = \max_x \min_y x^T Ay$.
How is this related to our above reasoning problem? Player 1 will be the user of the ontology, and player 2 will be the collection of all other agents in the system. The use of zero-sum games is convenient because it implicitly (and perhaps pessimistically) assumes the other agents to

cooperate with each other so that they can cause as much harm to player 1 as possible. Of course, this assumption is dramatic and most surely not correct, but as we seek to assure the quality of decisions against all scenarios, it turns out as a sharp worst-case scenario sketch. This is made explicit in the following result:

Proposition 4.1 (Rass & Schartner (2009)). *Let* $\Gamma = (N, PS, H)$ *with* $N = \{1,2\}$, $PS = \{PS_1, PS_2\}$, *and* $H = \{x^T Ay, x^T By\})$ *be a bi-matrix game with game-matrices* $A \in \mathbb{R}^{|PS_1|\times|PS_2|}, B \in \mathbb{R}^{|PS_2|\times|PS_1|}$ *for player* 1 *(honest) and player* 2 *(adversary), respectively. Let* $\Gamma_0 = (N, PS, \{x^T Ay, x^T(-A)y\})$ *be the zero-sum game from player 1's perspective (i.e. player 2 receives the payoff* $-x^T Ay$*), and let* $v(\Gamma_0)$ *denote its value (i.e. average outcome under a Nash-equilibrium strategy in* Γ_0*). Then*

$$v(\Gamma_0) \leq (x^*)^T Ay^* \tag{5}$$

for all Nash-equilibria (x^*, y^*) *of the game* Γ.

The proof is by simply observing that player 2 can either play the zero-sum strategy of Γ_0 (in this case the assumption is valid and we get equality in (5)) or act according to his own wishes. In the latter case, he necessarily deviates from the zero-sum strategy and thus increases the expected revenue for player 1.

4.2 Reasoning games

The observation that a zero-sum game soundly models a worst-case scenario from one player's point of view (proposition 4.1) leads to a simple way of assuring the quality of a decision: whenever we are facing random behavior, proposition 4.1 permits calculating the worst-case distribution and provides us with a behaviorial rule so that we get an assured outcome under this worst imaginable scenario. This is what we call

> *Assurance:* when facing an uncertain situation, our recommendation should be such that it provides a guaranteed outcome, independently of how much the observed behavior deviates from the assumptions under which a decision was made.

Proposition 4.1 is the key to do this, and the process is made rigorous after the following
Example: let us return to the introductory example sketched in section 1.1. We now invoke the game-theory and the (properly debugged) ontology to get the best answer from the three candidates.
Recall that the recommendations were $PS_1 = \{(s,f),(l,m),(o,h)\}$, where

(s,f): drive straight (s) at high speed (f)

(l,m): turn left (l) at the next junction, speed can be moderate (m).

(o,h): turn over (o) with high speed (h).

The oncoming traffic can either be slow or fast, and is free to turn left, right or straight at the next junction. Hence, the set PS_2 is composed from each of possible combinations, i.e. $PS_2 = \{\text{slow}, \text{fast}\} \times \{\text{turn left}, \text{turn right}, \text{go straight}\}$, making up 6 combinations, which we abbreviate as pairs in $PS_2 = \{(s,l),(f,l),(s,r),(f,r),(s,s),(f,s)\}$.
Assume that the ontology can decide upon the likelihood of an accident for each combination in $PS_1 \times PS_2$. For example, if the recommendation is to drive straight at high speed, and the oncoming traffic goes left, then the likelihood of an accident is higher than it would be if the oncoming traffic goes straight too (considering which driver has to give priority). If the

ontology classifies the accident likelihood in discrete terms like "none" (value 0), "negligible" (value 1) and "significant" (value 2), then the resulting matrix A could look like

A	(s,l)	(f,l)	(s,r)	(f,r)	(s,s)	(f,s)
(s,f)	1	2	0	0	0	0
(l,m)	0	0	1	2	1	2
(o,h)	0	0	0	0	1	2

Solving this game for its equilibrium value gives $v(A) = 0.5$ with equilibrium strategies $x^* = (1/2, 1/2, 0)$ and $y^* = (1/2, 0, 1/2, 0, 0, 0)$. Observe that this indicates that – aiming at maximal safety – we should never be advised to turn over, and can take either of the remaining choices with equal probability. Following this rule, we end up having a less than negligible chance (the value is 0.5 and as such strictly less than the negligible-value 1) of having an accident.

This process can be repeated for different scenarios, but crucially hinges on the ontology to be *correct* and perform reasoning *efficiently*.

Each query to the system yields a different matrix-game with strategy sets PS_1, PS_2, with its own unique Nash-equilibrium solution (which can as well be pre-computed). The actual recommendation provided to the user is a random selection from PS_1, where the particular choice is drawn from the equilibrium profile. This solution, among the valid alternatives, is presented as the recommendation, along with possible alternatives that are not explicitly recommended but possible.

This method is indeed computationally feasible, as a large set of possible games along with their corresponding Nash-equilibria can be *pre-computed* and stored for later usage in a *game-database* (indeed, all we need is the strategy set PS_1 and the corresponding Nash-equilibrium; the game matrix itself can be discarded). If the input parameters are discrete, then the set of queries is finite and hence the number of such games remains tractable. This is even more substantiated by the fact that a human operator will most likely not enter more than a few parameters as well as these will not be entered at arbitrary precision. As humans tend to reason in fuzzy terms, any natural mapping of these to parameters of a query answering system will consist of a small number of inputs to specify within small ranges to get an answer.

The ontology is then used to select the particular game at hand and provide the best behavior under uncertain behavior of others. The overall workflow is depicted in figure 3. The only block to be further explained in this picture is the *sampler*, drawing the concrete recommendation from the set of possible ones according to the Nash-equilibrium. This is a trivial task, as it amounts to sampling from a discrete probability distribution. We refer the interested reader to Gibbons (1992) for a formal justification, as we will restrict our presentation to giving the sampling algorithm: assume that a discrete distribution is given by $(p_1, \ldots, p_n)$.

1. Generate a uniform random number x within the interval $[0, 1]$.
2. Find and return (as the result) the smallest integer k such that $x \leq \sum_{i=1}^{k} p_i$.

So the overall process of reasoning with assurance can be described in a sequence of simple steps, presupposing a logically consistent ontology:

1. Unless the current query is found in the set of pre-computed ones,
2. generate the set of candidate recommendations,

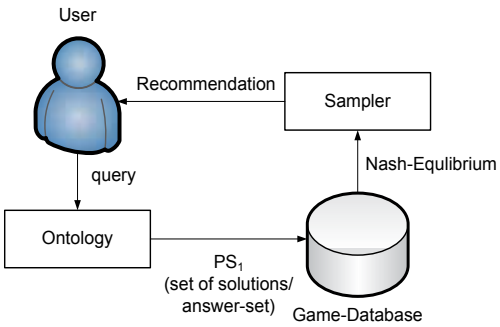


Fig. 3. Reasoning with game-theoretic assurance

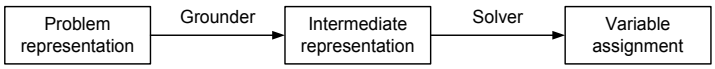


Fig. 4. Architecture of an ASP System (cf. Gebser (2008))

3. set up a game-matrix and solve for its equilibrium,
4. sample from the equilibrium profile and return the (so randomly selected) recommendation to the user.

From this process, it is evident that ontology and query optimization are major concerns, and therefore receive attention in the next section.

5. Maximizing reasoning performance

Optimization is the act of design and developing systems to take maximum advantage of the resources available. Ontology optimization formally defines the vocabularies in describing optimization methods and their configurations aiming to provide knowledge guidance in optimization selection and configuration Tao et al. (2004). Query optimization over an ontology consider the possible query plans to determine which of those plans will be most efficient. An ASP system usually consists of a grounder a grounder and a solver (see figure 4). First, a grounder translate A logic program in the language of AnsProlog s a non-ground problem description into a propositional program, which can be processed by the second component of the system, the solver (cf. Gebser (2008)).
There are algorithms that can be used for an effective optimization strategy for queries on knowledge databases query optimization.
One of the hardest problems in query optimization is the accurate estimation of the costs of alternative query plans. Cardinality estimation depend on estimates of the selection factor of predicates in the query. The estimation of the cardinality of a query usually is used to approximate the data transfer times of the result set, as part of the estimation of the total cost of executing a query (cf. Gebser (2008)). As an example a maximal complete query pattern path consists of a maximal query pattern path and a set of value constraints (cf. Shironoshita et al. (2007)). The total cardinality of such a path is obtained by calculating the product of the cardinality estimate of its maximal query path with all the value ratios for every variable in the query pattern path (cf. Shironoshita et al. (2007)).
In a highly distributed architecture where data in different locations connected through the internet, this is the most critical aspect of query execution time and the speed of the connections and the amount of data plays an important role.

A query optimization strategy is crucial to obtain reasonable performance over queries against ontology data models, especially if they are done over a highly distributed architecture.
However, using the answer set programming paradigm, we are able to model the spatial and temporal context models and define the relationships between classes and individuals (rules) and generating the answer sets using advanced solvers that support binder splitting, backjumping and all other features. The statements can be optimized to find the maximal or the minimal answer set of the logic program. The statement can be weighted or not, therefore weights can be omitted. If there are several minimize maximize statements in the logic program the latter will be preferred. The difference between the performance of ASP solvers and other existed ontology query languages is clearly high.The estimation of the cardinality of a query usually is used to approximate the data transfer times of the result set, as part of the estimation of the total cost of executing a query.

6. Conclusion

In the light of the growing importance and complexity of nowadays ontologies, tools that support debugging of reasoning systems are as well an increasing demand. In many cases, particularly for legacy systems, complex rules and facts that contribute to an inconsistency can present themselves as mere "black-boxes" to the human engineer. Fine-grained diagnosis targets at turning such black-boxes into white-boxes by enhancing the accuracy of the diagnosis up to the smallest grammatical construct that the language permits. Even if an ontology is consistent, the user should not be overloaded with (logically correct) answers, without a tool for choosing among them. Besides taking the degree of satisfaction as a selection criterion, the concept of assurance yields answers that are provably optimal according to a user-definable goal. Hence, we can have the ontology give us the answer that will maximize the benefit for the user. The process is cyclic in the sense that assurance hinges on consistency and vice versa. Moreover, retrieving answers timely calls for optimization, which in turn can make rules and facts even more complicated, again calling for fine-grained diagnostics. This chapter is hoped to provide useful starting points to enter this (implicit) cycle at any point for future research.

7. References

Baral, C., Gelfond, G., Son, T. & Pontelli, E. (2010), 'Using answer set programming to model multi-agent scenarios involving agents' knowledge about other's knowledge', pp. 259–266.

Brewka, G. (2002), 'Logic programming with ordered disjunction', pp. 100–105.

de Kleer, J. (1976), 'Local methods for localizing faults in electronic circuits', *MIT AI Memo* 394.

de Wolf, R. & Nienhuys-Cheng, S. H. (1997), *Foundations of inductive logic programming*, Springer Verlag, Berlin, Heidelberg, New York.

Eiter, T., Ianni, G., Schindlauer, R. & Tompits, H. (2005), 'Nonmonotonic description logic programs: Implementation and experiments', pp. 511–527.

Eiter, T. & Polleres, A. (2006), 'Towards automated integration of guess and check programs in answer set programming: a meta-interpreter and applications', *Theory and Practice of Logic Programming* 6(1-2), 23–60.

Friedrich, G., Rass, S. & Shchekotykhin, K. (2006), A general method for diagnosing axioms, *in* 'DX'06 - 17th International Workshop on Principles of Diagnosis', C.A. Gonzalez, T. Escobet, B. Pulido, Penaranda de Duero, Burgos, Spain, pp. 101–108.

Friedrich, G. & Shchekotykhin, K. (2005), A general diagnosis method for ontologies, *in* '4th ISWC', LNCS volume 3729, Springer.

Fuchs, S. (2008), *A komprehensive Knowledge Base for Context-Aware Tactical Driver Assistance Systems*, Shaker Verlag.

Garey, M. R. & Johnson, D. S. (1979), *Computers and intractability*, Freeman, New York.

Gebser, M, K. R. K. B. O. M. S. T. T. S. (2008), *Using gringo, clingo and iclingo*, UNI POTSDAM.

Ghose, D. & Prasad, U. R. (1989), 'Solution concepts in two-person multicriteria games', *Journal of Optimization Theory and Applications* 63(2), 167–189.

Gibbons, R. (1992), *A Primer in Game Theory*, Pearson Education Ltd.

Glicksberg, I. L. (1952), A further generalization of the Kakutani fixed point theorem, with application to nash equilibrium points, *in* 'Proceedings of the American Mathematical Society', Vol. 3, pp. 170–174.

Greiner, R., Smith, B. A. & Wilkerson, R. W. (1989), 'A correction to the algorithm in reiter's theory of diagnosis', *Artificial Intelligence* 41(1), 79–88.

Junker, U. (2004), QUICKXPLAIN: Preferred explanations and relaxations for over-contrained problems, *in* 'Proceedings of AAAI 04', San Jose, CA, USAs, pp. 167–172.

Kohler, J., Philippi, S. & Lange, M. (2003), 'SEMEDA: ontology based semantic integration of biological databases', *Bioinformatics* 19(18), 2420.

Matheus, C., Baclawski, K., Kokar, M. & Letkowski, J. (2005), 'Using SWRL and OWL to capture domain knowledge for a situation awareness application applied to a supply logistics scenario', pp. 130–144.

Rass, S. (2005), The diagnosis problem for axioms, Master's thesis, Universitaet Klagenfurt.

Rass, S. & Schartner, P. (2009), Game-theoretic security analysis of quantum networks, *in* 'Proceedings of the Third International Conference on Quantum, Nano and Micro Technologies', IEEE Computer Society, pp. 20–25.

Reiter, R. (1987), 'A theory of diagnosis from first prinicples', *Artificial Intelligence* 32(1), 57–95.

Schlobach, S. & Cornet, R. (2003), Non-standard reasoning services for the debugging of description logic terminologies, *in* 'Proceedings of IJCAI 03', Acapulco, Mexico, pp. 355–362.

Shironoshita, E., Ryan, M. & Kabuka, M. (2007), 'Cardinality estimation for the optimization of queries on ontologies', *ACM SIGMOD Record* 36(2), 13–18.

Snidaro, L., Belluz, M. & Foresti, G. (2008), 'Representing and recognizing complex events in surveillance applications', pp. 493–498.

Tao, F., Chen, L., Shadbolt, N., Xu, F., Cox, S., Puleston, C. & Goble, C. (2004), 'Semantic web based content enrichment and knowledge reuse in e-science', *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE* pp. 654–669.

Wang, X., Da Qing Zhang, T. & Pung, H. (2004), 'Ontology based context modeling and reasoning using OWL'.

World Wide Web Consortium (2010), 'Resource description framework (RDF)', `http://www.w3.org/RDF/`. (last accessed: 28th January, 2010).

6

Common Sense Reasoning in Diagnostic Systems

Alexander P. Eremeev and Vadim N. Vagin
Moscow Power Engineering Institute (Technical University)
Russia

1. Introduction

The problem of human reasoning modeling (so called "common sense" reasoning) in artificial intelligence systems and especially in *intelligent decision support systems (IDSS)* is very actual nowadays [Vagin et al., 2001]. That is why special attention is turned to case-based and analogous reasoning methods and models. The analogies and precedents (cases) can be used in various applications of artificial intelligence (AI) and for solving various problems, e.g., for diagnostics and forecasting or for machine learning. AI experts model case-based reasoning by computers in order to develop more flexible models of search for solutions and learning.

Investigation of mechanisms that are involved in the analogous reasoning process is an important problem for the specialists in AI. The analogy can be used in various applications of AI and for solving various problems, e.g., for generation of hypotheses about an unknown problem domain or for generalizing experience in the form of an abstract scheme. The great interest in this problem is caused by the necessity of modeling human reasoning (common sense reasoning) in AI systems and, in particular, in IDSS of real time.

Reasoning by analogy is to transfer of knowledge obtained from an object to a less studied one which is similar to the former with respect to some essential properties or attributes. Reasoning of this kind is a source of scientific hypotheses. Thus, analogy-based reasoning can be defined as a method that allows to understand a situation when compared with another one. In other words, an analogy is an inference method that allows to detect likeness between several given objects due to transfer of facts and knowledge valid for both objects, to other objects and to determine means of problem solution or to forecast unknown properties.

Case-based reasoning, like reasoning by analogy, is based on analogy; however, there are certain differences in their implementation. In the most encyclopedias, a precedent (from Latin, precedentis) is defined as a case that took place earlier and is an example or justification for subsequent events of this kind. Creating a precedent is to give grounds for similar cases in the future, and establishing a precedent is to find a similar case in the past.

The generalized structure of a real-time IDSS (RT IDSS) is represented in Fig. 1 [Vagin et al., 2007].

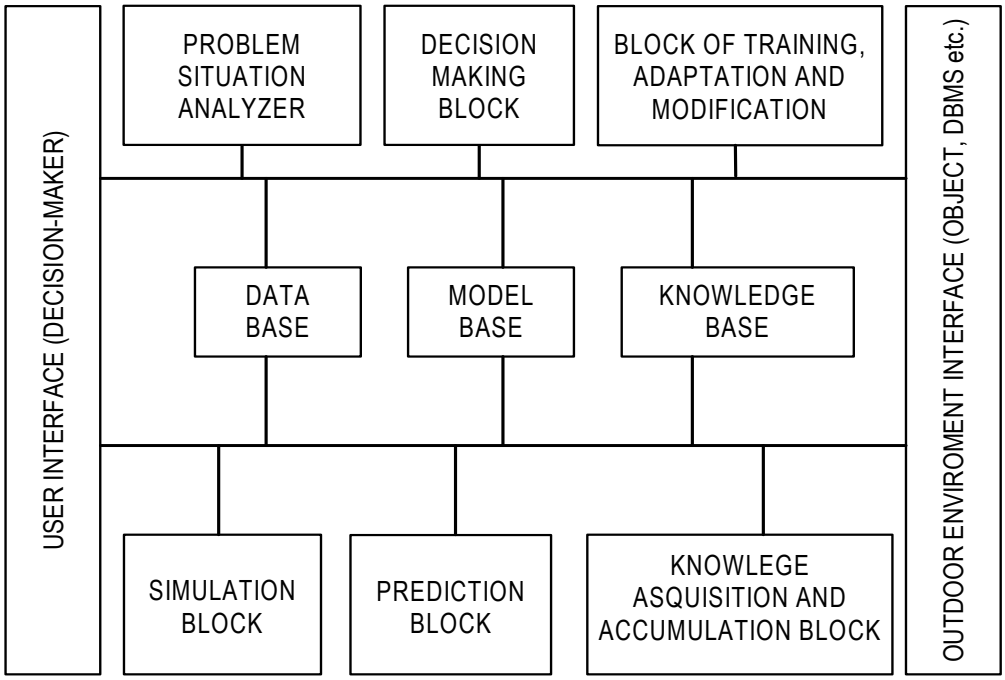


Fig. 1. The generalized structure of a real-time IDSS

Formally, a RT IDSS can be defined by the set
SS = <M, R(M), F(M), F(SS)>,
where
$M = \{M_1,\ldots, M_n\}$ is the set of formal or logic-linguistic models, implementing defined intelligent functions;
$R(M)$ is the function for selection of the necessary model in a current situation;
$F(M) = \{F(M_1),\ldots, F(M_n)\}$ is the set of modification functions of models $M_1,\ldots, M_n$;
$F(SS)$ is the function for the SS modification, i.e. its basic components $M, R(M), F(M)$.
The main problems, solved by RT DSS, are:

- diagnostics and monitoring – revealing of problem situations;
- decision search – searching an optimal or admissible sequence of actions allowing to achieve the desired goal or solve the problem situations;
- forecasting – assessing the recommended actions related to the goal achievement (to solve a problem situation).

2. Reasoning by analogy

Nowdays there are a great number of various models, schemes, and methods that describe mechanisms of *reasoning by analogy* [Haraguchi et al., 1986; Long et al., 1994; Varshavskii et al., 2005; Eremeev et al., 2005].
In (Haraguchi et al., 1986), the authors have proposed two types of analogies, *an analogy for solving problems and an analogy for forecasting:*

- *The analogy for solving problems* assumes the application of reasoning by analogy for increasing the efficiency of the problem solution which, generally speaking, can be solved without analogy as well as e.g., in programming and proving theorems;
- *The analogy for prediction (forecasting)* uses reasoning by analogy for obtaining new facts. Due to the transformation of knowledge based on the likeness of objects, one can make the conclusion that new facts probably hold.

Depending on the nature of information transferred from an object of analogy to the other one, *the analogy of properties and the analogy of relations can be distinguished:*

- *The analogy of properties* considers two single objects or a pair of sets (classes) of homogeneous objects, and the transferred attributes are the properties of these objects, for example, analogy between illness symptoms of two persons or analogy in the structure of the surfaces of Earth and Mars, etc.;
- *The analogy of relations* considers pairs of objects where the objects can be absolutely different and the transferred attributes are properties of these relations. For example, using the analogy of relations, bionics studies processes in nature in order to use the obtained knowledge in a modern technology.

We consider the methods of solution search on the basis of structural analogy which allows to take into account a context and based on the theory of structural mapping. We use semantic networks as a model of knowledge representation.

2.1 Reasoning by structural analogy taking into account the context

Consider an **analogy** as a quadruple $A = \langle O, C, R, p \rangle$, where O and R are the source object and the receiver object and C is the intersection object, i.e., the object that structurally is intersected with the source object and receiver object, and has a larger cardinality of the set of properties in the comparison with these objects. In other words, the analogy between the source object and receiver object is considered in the context of the intersection C, and p is a property for the definition of an original context. The structure of this analogy is represented in Fig. 2.

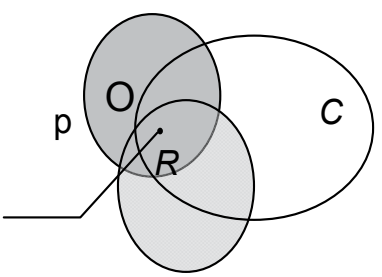


Fig. 2. Structure of analogy using the context

We use semantic networks (SNs) as a model of the knowledge representation for reasoning by analogy. The choice of an SN for the knowledge representation possesses an important advantage, which distinguishes it from other models, such as natural representation of structural information and fairly simple updating in a relatively homogenous environment. The latter property is very important for real-time IDSS oriented towards open and dynamical problem domains.

A semantic network is a graph **<*V*, *E*>** with labeled nodes and arcs, where ***V*** and ***E*** are sets of nodes and arcs, respectively. The nodes can represent objects (concepts, events, actions, etc.) of a problem domain, and the arcs represent relations between them.
By ***Pv*** , we denote the set of properties of an object $v \in V$.
Objects $v, v' \in V$ *intersect* each other on SN if and only if $Pvv' = Pv \cap Pv' \neq \varnothing$, where Pvv' is a set of common properties of objects v and v'.
By ***Vp*** , we denote a set of SN objects that have a property ***p***.
By ***Vv*** ,$Vv \subseteq V$, we denote an object set of objects that intersect $v \in V$.
The object ***C*** is *an intersection for* ***A*** if and only if there is ($C \in V$) & ($p \in PC$) & ($nR \leq nC$) & $\neg$($nR << nC$) & ($nRC < nR$) & ($nRC > 1$), where nR and nC are the numbers of properties of the receiver R and the intersection C, respectively; nRC is the number of their common properties, $\neg$($nR << nC$) denotes that receiver R should not be much smaller than intersection C (i.e., the possibility of absorbing the receiver R by the intersection C, since here the probability of receiving a false analogy increases).
The object ***O*** is *the source for analogy* ***A*** if and only if there is ($O \in V$) & ($p \in PO$) & ($nO \leq nC$) & $\neg$($nO << nC$) & ($nOC < nO$) & ($nOC > 1$), where nO is the number of properties of the source O; nOC is the number of common properties of the source O and intersection C; and other notations are analogous to the previous definition.
By ***VC***, $VC \subseteq Vp$, we denote the set of objects that are candidates for the role of intersection C for analogy A.
By $VO \subseteq Vp$, we denote the set of objects that are candidates for the role of source O for analogy A.
By ***VA***, we denote the set of analogies A.
The set $POCR = PO \cap PC \cap PR$ denotes *the context*, with respect to which analogy ***A*** is considered.
We consider the structure of the SN in detail (for Metalevel and for Situation 1) using the example from power engineering - operation control of the nuclear power unit (Fig. 3) [Eremeev et al., 2006a].

Let us give a semantic interpretation of the information given in the SN for Situation 1:

- It is recommended to supply the pump TH11D01 with boric concentrate 40g/kg caused by switching off automatic cooling system ACS 1 due to closing the gates TH11S24 and TH11S25;
- ACS 2 is switched off due to the closed gates TH12S24 and TH12S25;
- The upper setting T517B01 is equal to 63;
- The lower setting T517B01 is equal to 56;
- The upper setting TH11T500 is equal to 60;
- The lower setting TH11T500 is equal to 20.

Analogously, the fragments of the SNs, which illustrates Situations 2,3,4, are represented in Fig. 4.

2.2 Algorithm of reasoning by structural analogy

An SN with information about the problem domain, a receiver R, and the property for defining the original context ***p*** provide input data for this algorithm.
The algorithm for the problem solution on the basis of analogy taking into account the context consists of the following steps:

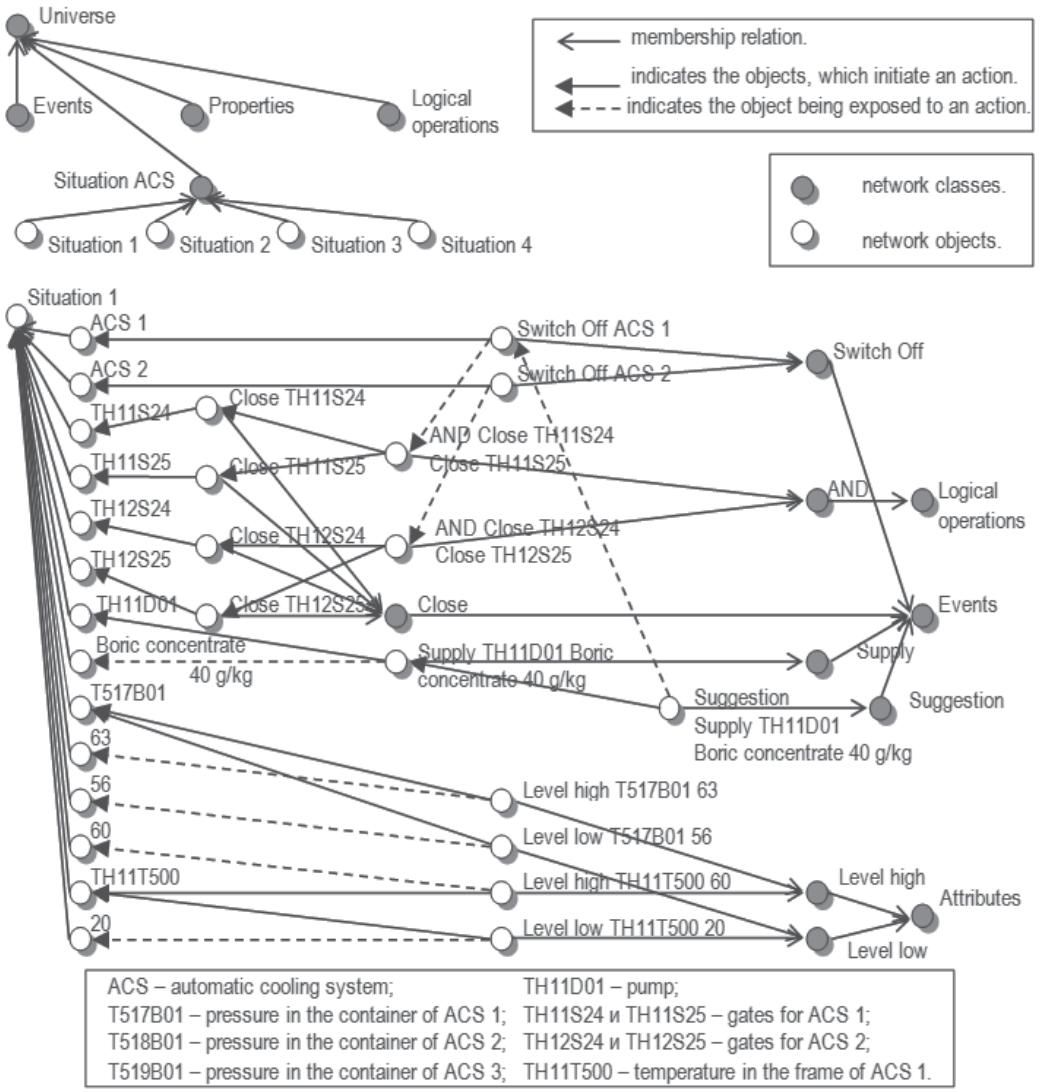


Fig. 3. A fragment of the SN that represents the Metalevel and the Situation 1 that was formed in the course of ACS functioning

Step 1. $VC = \varnothing$, $VO = \varnothing$, $VA = \varnothing$. Determine all objects of the SN, except for receiver R, that have property p ($Vp' = Vp \setminus \{R\}$). If there are no objects of this kind, then the search for a solution fails (without finding an analogy), otherwise, go to **step 2**.

Step 2. For the objects found in step 1, determine all possible intersections of C with R taking into account p (VC). If there are no intersections of C with R ($VC=\varnothing$), the first search for a solution fails, otherwise, go to step 3.

Step 3. From the objects extracted in step 1, determine all possible sources O for analogies (VO). In the case of success ($VO \neq \varnothing$), go to step 4, otherwise, the search for a solution fails.

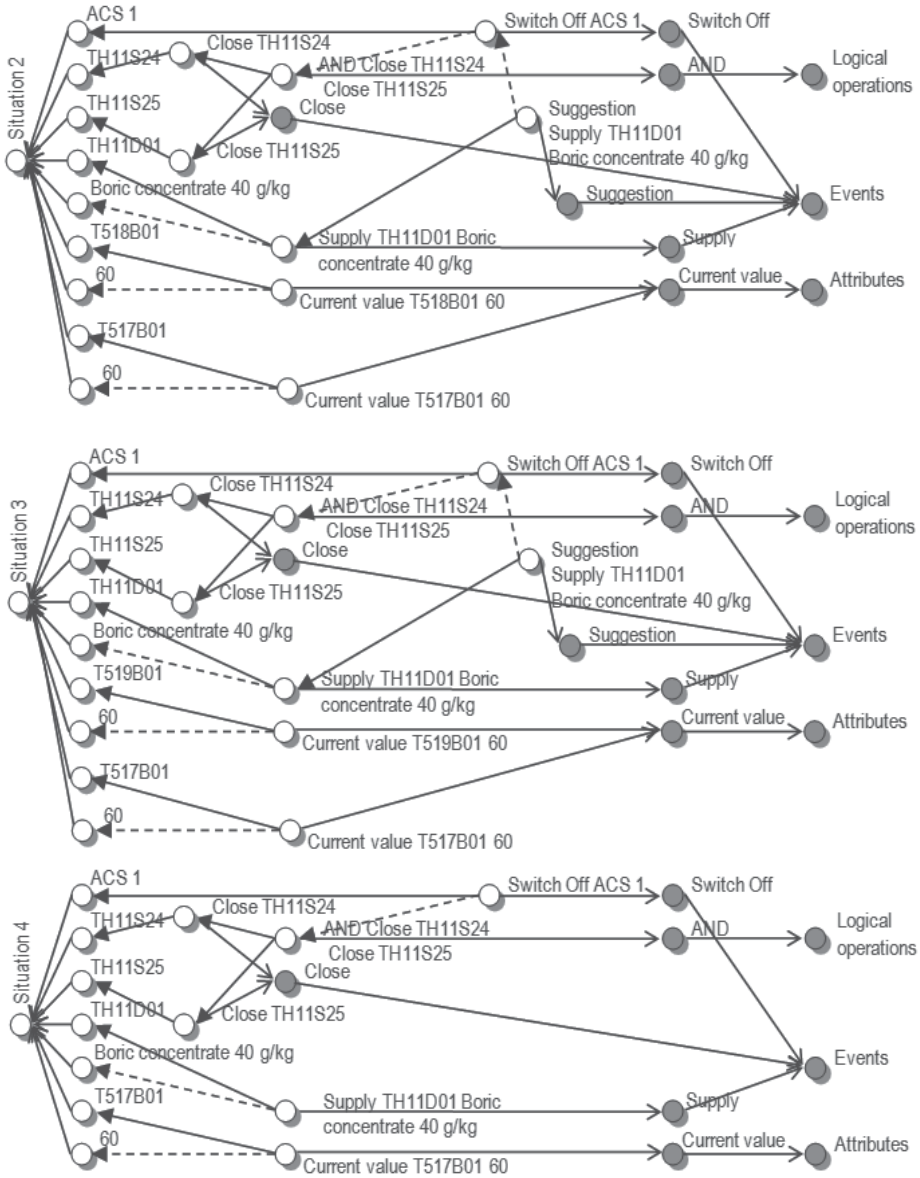


Fig. 4. The fragments of the SNs that represent Situations 2-4

Step 4. Construct possible analogies for ***R*** using the sets ***VC*** and ***VO***. Add new analogy ***A*=⟨*O*,*C*,*R*,*p*⟩** to ***VA*** if and only if there exists an analogy ***A*'=⟨*O*',*C*,*R*,*p*⟩**, $O \neq O'$. In the case of success ($VA \neq \varnothing$), go to step 5; otherwise, the search for a solution fails.

Step 5. The analogies obtained in step 4 (***VA***) (which could be previously compared with each other taking into account the context) are given to the decision making person (DMP), which means successful termination of the algorithm.

Having obtained analogies, the DMP may then make the final choice of the best ones. On the basis of these facts, the facts (properties) that hold for the source O are transferred to the receiver R.

Let us consider the steps of the functioning of the algorithm using the example from power engineering - operation control of the nuclear power unit (see Fig. 3,4).

As a receiver *R* for the analogy, we take Situation 4 and as the property *p*, we take Close TH11S24.

In the first step, $VC = \varnothing$, $VO = \varnothing$, $VA = \varnothing$ and Vp' = {Situation 1, Situation 2, Situation 3}. Since $Vp' \neq \varnothing$, we go to the next step.

Determine intersections of *C* with *R* taking into account *p*. Add in *VC* only Situation 1, because the number of common properties $nRC = nR$ for Situation 2 and Situation 3. Since $VC \neq \varnothing$, we go to the step 3.

Determine all possible sources *O* and go to step 4. In this case *VO* = {Situation 2, Situation 3}, because the Situation 1 is unique intersection for analogy.

In the fourth step, we construct only two analogies for *R* - Situation 4:

***A1* = <Situation 2, Situation *1*, Situation *4*, Close TH11S24 >;**

***A2* = <Situation *3*, Situation *1*, Situation *4*, Close TH11S24 >.**

Add new analogies to ***VA*** and go to step 5.

The analogies obtained in step 4 (*A1, A2*) are given to the DMP.

As a result we obtain two analogies. Choosing one of them, the DMP can transfer facts that hold for the source of the analogy to its receiver. In this example, a new fact about the recommendation *"Supply the pump TH11D01 with boric concentrate 40g/kg caused by switching off ACS 1 due to closing the gates TH11S24 and TH11S25"* arises for Situation 4.

The methods of reasoning by analogy is more general than on the bases of cases. Analogies are used when it is impossible to find a suitable case in a case library. The reasoning by analogy method can be used independently from a case- based reasoning method as well as for correction (adaptation) of the nearest to a problem situation case to form a new case for completing a case library. Further we shall consider the case-based reasoning method and its application.

3. Case-based reasoning

Case-based reasoning (CBR), like analogous reasoning, is based on analogy; however, there are certain differences in their implementation [Aamodt, 1994; Eremeev et al., 2006b].

As the practice shows, when a new problem situation arises, it is reasonable to use the method of case-based reasoning without drawing an analogy. This is caused by the fact that humans operate with these reasoning schemes at the first stages, when they encounter a new unknown problem.

Case-based reasoning is an approach that allows one to solve a new problem using or adapting a solution of a similar well-known problem. As a rule, case-based reasoning methods include four main stages that form a CBR-cycle, the structure of which is represented in Fig. 5.

The main stages of CBR-cycle as follows:

- Retrieving the closest (most similar) case (or cases) for the situation from the case library;
- Using the retrieved case (precedent) for solving the current problem;
- If necessary, reconsidering and adaptation of the obtained result in accordance with the current problem;

- Saving the newly made solution as part of a new case.

It is necessary to take into account that a solution on the basis of cases may not attain the goal for the current situation, e.g., in the absence of a similar (analogous) case in the case library. This problem can be solved if one presupposes in the CBR-cycle the possibility to update the case library in the reasoning process (inference). A more powerful (in detecting new facts or new information) method of reasoning by analogy is a method of updating case libraries. We also note that the elements of case-based reasoning may be used successfully in analogy-based reasoning methods; i.e., these methods successfully complement each other and their integration in IDSS is very promising.

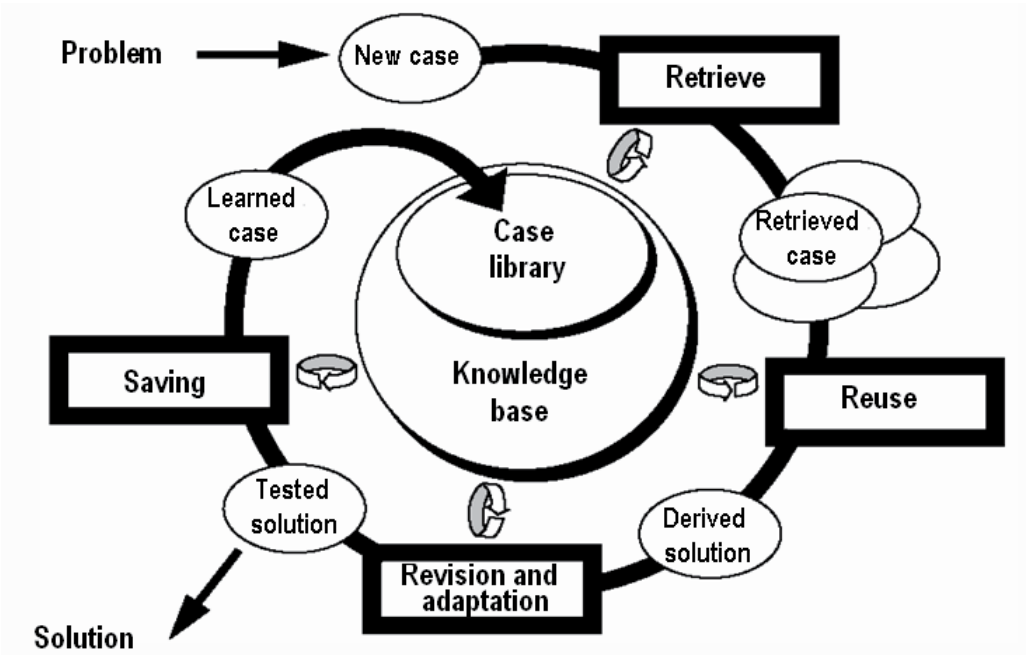


Fig. 5. The structure of CBR-cycle

Using the mechanism of cases for RT IDSS consists in issuing the decision to the operator (DMP – Decision Making Person) for the current situation on the basis of cases which are contained in a system. As a rule, the last stage in a CBR-cycle is excluded and performed by the expert (DMP) because the case library should contain only reliable information confirmed by the expert. Reconsidering and adaptation of the taken decision is required seldom because the same object (subsystem) is considered.

The modified CBR-cycle for RT IDSS includes following stages:

- Retrieving the closest (most similar) case (or cases) for the situation from the case library;
- Using the retrieved case (precedent) for solving the current problem.
- Case-based reasoning for IDSS consists in definition of similarity degree of the current situation with cases from a case library. For definition of similarity degree, the nearest neighbor algorithm (k-nearest neighbor algorithm) is used [Eremeev et al., 2007a].

Further, we shall view the structure of a case library for RT IDSS on the basis of non-classical logics for monitoring and control of complex objects like power units.
The case library for RT IDSS should join in itself the cases concerning a particular subsystem of a complex object, and also contain the information on each parameter which is used for the description of cases (parameter type and range). Besides, the case library should include such adjustments, as:

- the significance of parameter;
- a threshold value of similarity;
- a value which limits quantity of considered cases.

It is necessary to emphasize, that the case library can be formed on the basis of:

- the experience, accumulated by the expert;
- the analysis of the system archive;
- the analysis of emergencies;
- operative instructions;
- technological requirements.

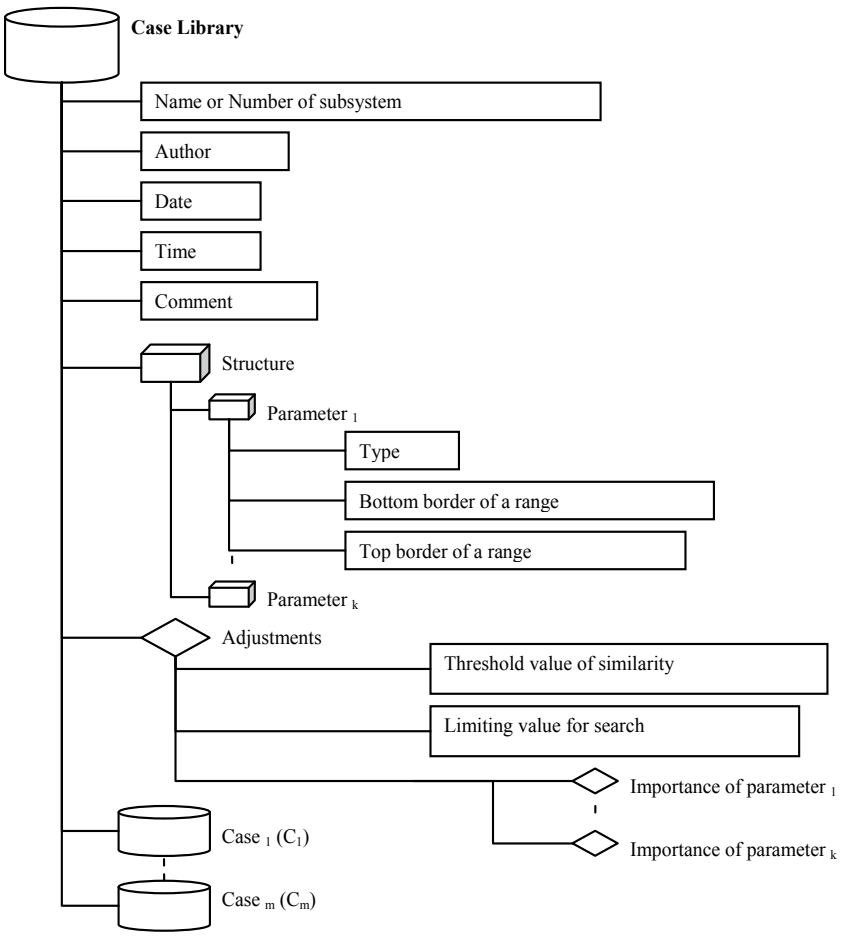


Fig. 6. The structure of the case library

The case library can be included in the structure of the knowledge base of RT IDSS or act as a separate component of the system. The structure of case library is represented in Fig. 6.

4. Application of case-based reasoning for diagnostics of complex object states

As a complex object, we shall view an object which has a complex architecture with various interrelations, a lot of controllable and operated parameters and small time for acceptance of operating actions. As a rule, such complex objects as the power unit are subdivided into technological subsystems and can function in various modes (regular, emergency, etc.).
For the description of such complex object and its subsystems, the set of parameters is used. The state of object is characterized by a set of concrete values of parameters.
In the operative mode, reading of parameters values from sensors for all objects is made by the system of controllers with an interval in 4 seconds. For this time interval, it is necessary to give out to the DMP (operator) the diagnosis and the recommendations on the given situation.
Diagnosing and detection of operating actions is carried out on the basis of expert knowledge, technological requirements and operative instructions. The developed software (Case Libraries Constructor – CLC) can be applied to the decision of the specified problems.
Basic components of CLC are:

- a module for storage and loading the cases to library and for data import;
- a subsystem of visualization for browsing the structure of a case library;
- a subsystem of editing and adjustment of a case library;
- a module of new case check;
- a subsystem of case library testing and case-based reasoning.

CLC was implemented in Borland C++ Builder 6.0 for Windows 7/ NT/XP.
This tool was applied in the prototype of a RT IDSS for monitoring and control of complex objects like power units on the example of a pressurizer in pressurized water reactor (PWR) of the atomic power station (Fig. 7) [Eremeev et al., 2007b, 2008].
Implementation of the case library with using CLC for expert diagnostics systems is subdivided into the following main stages:

- Creation of case libraries for subsystems of complex object;
- Adjustment of the created case libraries;
- Addition of cases in case libraries;
- Check of the added cases;
- Testing of the filled case libraries with using case-based reasoning;
- Reservation of the created case libraries for their subsequent transfer to operative maintenance.

5. Model-based diagnostics

Many diagnostics problems require building the behaviour prognoses, the work with contradictions and defaults, effective treatment of new facts and assumptions. The typical problem of diagnostics is to find a faults of a diagnosed device on the basis of some set of observations.
The generalized problem of diagnostic can be formulated as follows. There is a device exhibiting an incorrect behaviour. The device consists of components, one or several of

which are not working properly what is the reason of incorrect behaviour. There is a structure of connections between components and a possibility to get measurements on their inputs and outputs. It is necessary to determine what of components are faulty with minimal resource expenses. At present two main approaches to a solution of the given problem are viewed [Clansey, 1985; de Kleer et al., 1987; Forbus et al.,1993].
The first approach is heuristic diagnostics. The base of this approach is the knowledge extraction from an expert and building fault determining rules in the form of "symptoms → faults" [Clansey, 1985].

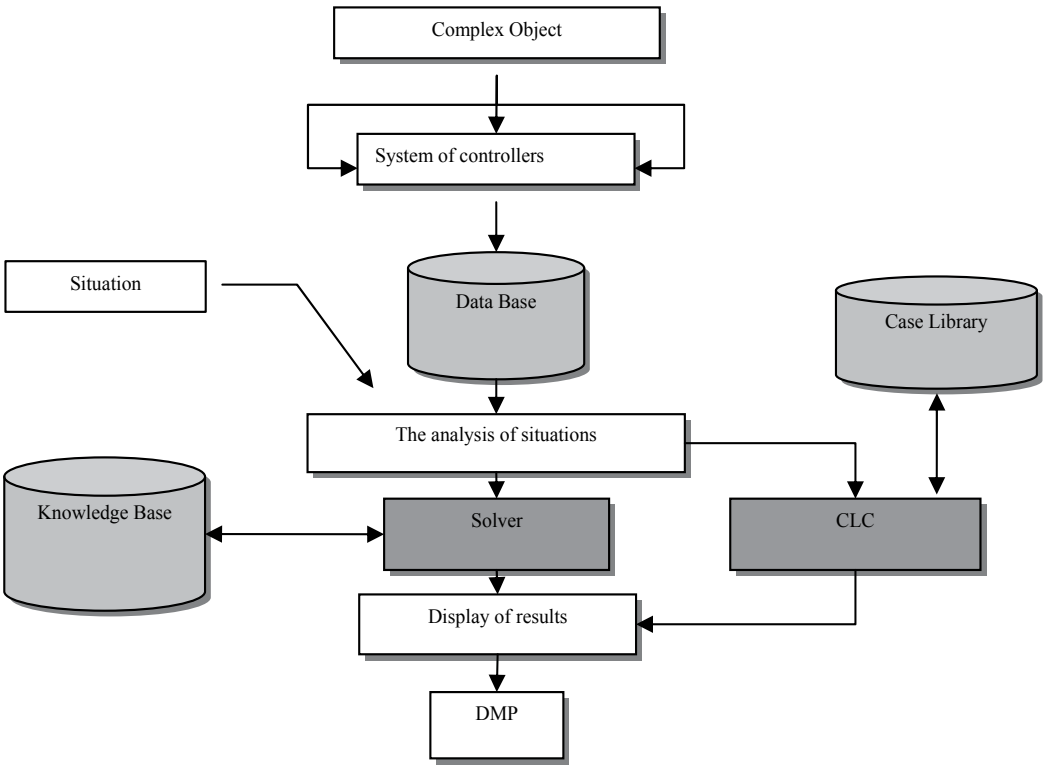


Fig. 7. The scheme of functioning for RT IDSS with use of CLC

Drawbacks of the given approach are:
- a complexity of multiple fault diagnostics;
- the problems in unexpected faults recognition;
- a complexity of expert knowledge formalizing;
- difficulties using the knowledge bases for other diagnostic problems;
- a rigid dependence from a device structure;
- a difficulty of explanations.

The second approach is model-based diagnostics [de Kleer et al., 1987; Forbus,1993]. This approach is based on the knowledge of device component functionality.
The model of a device is a description of its physical structure, plus the models for each of its components. A compound component is a generalized notion including simple components, processes and even logical inference stages.

Model-based diagnosis process is the comparison of predicted device behaviour with its observed behaviour (Fig. 8).
It is supposed, that the model is correct, and all differences between device behaviour and a device model indicate availability of broken components.
Main advantages of the model-based approach:

- diagnosing the multiple faults;
- unexpected fault recognition;
- a precision of a component model description does not depend on the expert experience;
- a possibility of new device diagnosing;
- multiple using the models;
  - detailed explanations.

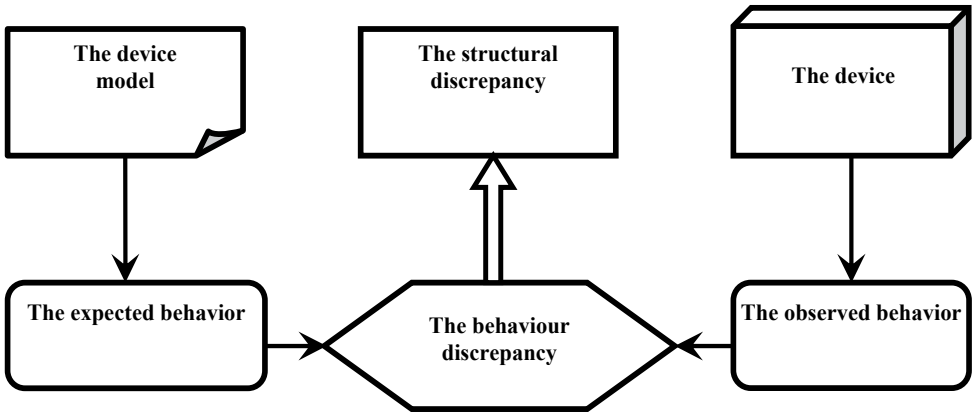


Fig. 8. Model-based diagnosis process

6. Assumption-based truth maintenance systems

For building a prognosis network, a component behaviour model, finding minimal conflicts characterizing irrelevance of observations with prognoses and minimal candidates for a faulty, it is profitable to use possibilities given by Assumption-based Truth Maintenance Systems (ATMS) [de Kleer, 1986; Vagin et al., 2008].
The Truth Maintenance Systems (TMS) are the systems dealing with the support of a coherence in databases. They save the assertions transmitted to them by a problem solver and are responsible for maintaining their consistency. Each assertion has the justification describing what kind of premises and assumptions this justification was obtained. The environment is a set of assumptions.
The inference of an inconsistency characterizes assumption incompatibility within the presuppositions of which this conclusion was made. Also there is introduced the environment set which contains some inconsistency [de Kleer, 1986]. The sets of inconsistency environments $E_1,E_2,...,E_m$ are Nogood = $\{E_1,E_2,\ldots,E_m\}$. A consistent ATMS environment is not Nogood.
There are the following correspondences between ATMS and the model-based diagnosis approach:

- ATMS premises - an observed device behaviour;
- ATMS assumptions - components of a device;
- inferred ATMS nodes - predictions of an diagnostic system;
- Nogood - the difference between predicted and observed device behaviour.

7. The current measurement point determination

One of the key aspects of the model-based fault search algorithm is to determine the optimal current measurement in a diagnosed device [de Kleer, 1987]. Efficiency of the current measurement choosing allows essentially reducing a decision search space while the inefficiency of choice will increase an operating time, the space of a searching algorithm, and also require additional resource spends to implement a measurement.
The best measurement point in a diagnosed device is a place (point) of measuring a value giving the largest information promoting the detection of a set of fault components at minimal resource spending.
One of the best procedures for reducing resource expenses is to produce the measuring giving the maximal information concerning predictions made on the basis of the current information on a system.

7.1 Heuristic methods of choosing a measurement point

The purpose of the best choosing a measurement point is to derive the maximal component state information. After each measuring there is a confirmation or refutation of prediction values in a point of measurement. So, it is possible to use the following aspects [Vagin et al., 2006a,b,c]:

- knowledge about environments that support predicted values in the measurement points which can be confirmed or refuted;
- knowledge about inconsistent environments;
- knowledge about coincided assumptions of the inconsistent environments.

7.2 Knowledge about supporting environments

The diagnostic procedure constructs predictions of values for each device point with the list of environments in which the given prediction is held. The list of environments represents assumption sets about correctness of corresponding device components.
Let's consider a sample of the 3-bit parity checker (Fig. 9) [Frohlich, 1998].
The vector (1, 0, 1) is input to device components (Inv1, Inv2, Inv3). Let the
following measurements be obtained: D=0, R4=0, R2=1. The mismatch between observations and predictions speaks about a fault in a device. Based on measured observations additional predictions of values are formed. In general, it is obtained the following set of predictions with appropriate environments:
<R3=0, {{And3,Inv1}, {And2,And3,Inv6}, {And2,And3,Inv5}}>;
<R1=0, {{And1,Inv3}, {And1,And2,Inv4}, {And1,And2,Inv5}}>;
<O6=0, {{And2,Inv6}}>;
<O6=1, {{Inv3,Inv6}}>;
<O5=1, {{And2}}>;
<O5=0,{{Inv2,Inv5}, {And4,Inv1,Inv3,Inv4,Inv6}}>;
<O4=0, {{And2,Inv4}}>;

<O4=1, {{Inv1,Inv4}}>;
<O3=1, {{And2}}>;
<O3=0, {{Inv3}}>;
<O2=0, {{And2,Inv5}}>;
<O2=1, {{Inv2}, {And4,Inv1,Inv3,Inv4,Inv5,Inv6}}>;
<O1=1, {{And2}}>;
<O1=0, {{Inv1}}>.

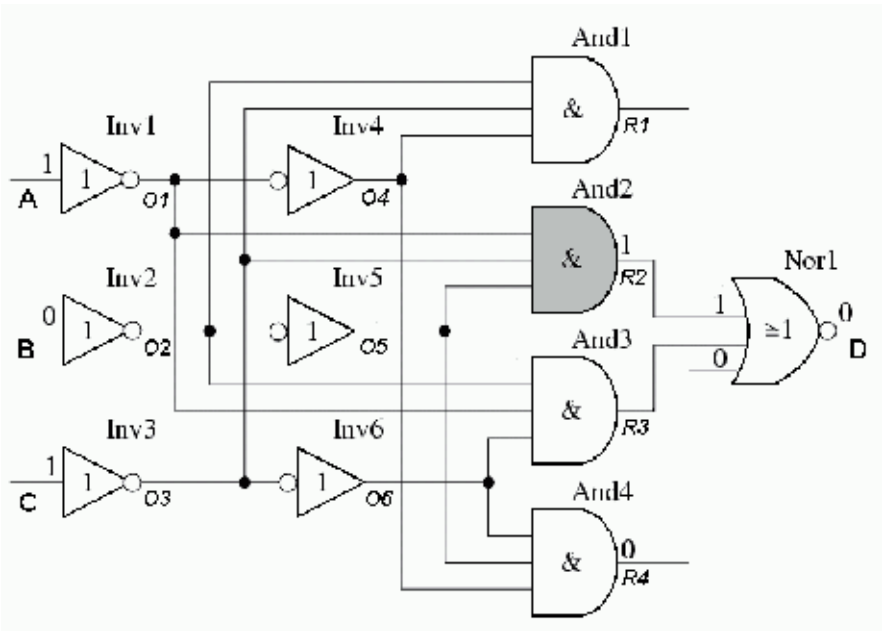


Fig. 9. The 3-bit parity checker

As we are interested with a measurement point with the greatest information on failure the point is selected from a quantity of assumptions.
Designate an environment set as *Envs*(*x*). So, for the considered example, *Envs*(*O1*)={{And2}, {Inv1}}. Let's introduce the function *Quan*(*x*), by which we will designate the information quantity obtained at measuring values in the point *x*.
If the environment *J* represents a unique assumption, then obviously the set cardinality will be equal 1: $|J| = 1$. The information quantity obtained from such environment is equal to 1. If the environment consists more than one component the information quantity obtained at confirming or refuting a value is less because we have knowledge not about a concrete valid / fault component but about a component set among of which are faulty. Therefore the information quantity obtained from an environment consisting of more than one assumption, we heuristically accept equal to half of set cardinality. Thus the function *Quan*(*x*) is:

$$Quan(x) = \sum_{\substack{J_i \in Envs(x) \\ |J_i|=1}} |J_i| + \sum_{\substack{J_j \in Envs(x) \\ |J_j|>1}} \frac{|J_j|}{2} \tag{1}$$

Summing is produced on all possible values in the point x.
Then for the presented list of considered example predictions there are:
Quan(R3) = (2/2 + 3/2 + 3/2) = 4
Quan(R1) = (2/2 + 3/2 + 3/2) = 4
Quan(O6) = (2/2) + (2/2) = 2
Quan(O5) = (1) + (2/2 + 5/2) = 4.5
Quan(O4) = (2/2) + (2/2) = 2
Quan(O3) = (1) + (1) = 2
Quan(O2) = (2/2) + (1 + 6/2) = 5
Quan(O1) = (1) + (1) = 2
Points with the greatest value of the function $Quan(x)$ have the greatest priority of a choice. We will call the given method of choosing a measurement point as SEH (Supporting Environment Heuristics).
In our example the point giving the greatest information quantity is O2.

7.3 Knowledge about the sets of inconsistent environment

As a result of each measurement there is a confirmation or refutation of some prediction. The environments $E_1,E_2,...,E_m$ of refuted prediction form the set $Nogood = \{E_1, E_2,...,E_m\}$. It can be used for directional searching for more precise definition what kind of components from *Nogood* is broken.
Let we have the set *Nogood* = {{*And2*,*Invl*}, {*And2*,*Inv2*,*Inv5*}, {*And2*,*Inv3*}}.
Obviously the more of the components from *Nogood* are specified by measuring a value in some device point, the more the information about which components of *Nogood* are broken will be obtained. For using this possibility, it is necessary to take the intersection of each environment from *Envs(x)* with each set from *Nogood*:

$$Envs(x) \cap Nogood = \{A \cap B : A \in Envs(x), B \in Nogood\}.$$

Continuing the example at Fig. 9, we obtain the following:

<R3, {{Invl}, {And2}, {And2,Inv5}}>

<R1, {{Inv3}, {And2}, {And2,Inv5}}>

<06, {{And2}, {Inv3}}>

<05, {{And2}, {Inv2,Inv5}, {Invl}, {Inv3}}>

<04, {{And2},{Invl}}>

<03, {{And2},{Inv3}}>

<02, {{And2}, {And2,Inv5}, {Inv2}, {Invl}, {Inv5}, {Inv3}}>

<01, {{And2}, {Invl}}>

For this approach the equation (1) can be changed as follows:

$$QuanN(x) = \sum_{\substack{J_i \in Envs(x) \cap Nogood \\ |J_i|=1}} |J_i| + \sum_{\substack{J_j \in Envs(x) \cap Nogood \\ |J_j|>1}} \frac{|J_j|}{2}$$

For each of these points we calculate the information quantity as follows:
QuanN(R3) =1 + 1 + 2/2 = 3
QuanN(Rl) =1 + 1+2/2 = 3
QuanN(06) =1 + 1=2
QuanN(05) =1+2/2+1 + 1=4
QuanN(04) =1 + 1=2
QuanN(03) = 1 + 1 = 2
QuanN(02) = 1 + 2/2+1 + 1 + 1 + 1 = 6
QuanN(Ol) = 1 + 1=2
Points with the greatest value of function QuanN(x) have the greatest priority of a choice. We will call the given method of choosing a measuring point as SIEH (Supporting and Inconsistent Environment Heuristics).
One can see that the point 02 as the most informative is again offered. And in the given approach the difference in information quantity between various points is expressed more brightly than without Nogood usage.

7.4 Knowledge about coincided assumptions of the inconsistent environments

During diagnostics of faulty devices as a result of confirmations and refutations of some predictions there is a modification of a set of inconsistent environments Nogood.
In each component set from Nogood one or more components are broken what was a reason of including a supporting set into the inconsistent environments Nogood. Taking the intersection of all sets of the inconsistent environments, we receive a set of components which enter into each of them, so their fault can be a reason explaining an inconsistence of each set holding in Nogood. Thus, we obtain the list of components a state of which is recommended to test first of all, i.e. the most probable candidates on faultiness.
The set intersection of inconsistent environments is expressed by the following equation:

$$SingleNogood = \bigcap_{E_i \in Nogood} E_i$$

For Nogood = {{And2, Invl}, {And2, Inv2, Inv5}, {And2, Inv3}} the set of the most probable candidates will be the following: SingleNogood = {And2}.
If SingleNogood = ∅, it means that there are some disconnected faults. In this case the given approach is inapplicable and it is necessary to define more precisely the further information by any other methods.
After obtaining a set SingleNogood ≠ ∅, on the base of environments of value predictions in device points it is necessary to select those measurement points that allow to effectively test components to be faulted from SingleNogood.
For this purpose we will work with the sets obtained as a result of an intersection of each environment from *Envs(x)* with *SingleNogood:*

$$\text{Envs(x)} \cap \text{SingleNogood} = \{\text{J} \cap \text{SingleNogood} : \text{J} \in \text{Envs\{x)}\}$$

The following versions are possible:

a. $\exists\, J \in Envs(x)$: $J \equiv SingleNogood$. One of environments of the value prediction in the point x coincides with the set SingleNogood. The given version allows to test faulty components from the set SingleNogood most effectively so this measurement point *x* is selected with the most priority.

b. $\exists J \in$ Envs(x): |J ∩ SingleNogood| < |SingleNogood|. The cardinality of SingleNogood is more than the cardinality of a set obtaining as result of an intersection SingleNogood with a set from Envs(x). We evaluate this version as $\max_{J \in Envs(x)} |J \cap SingleNogood|$ i.e. the more of components from SingleNogood are intersected with any environment from Envs(x), the more priority of a choice of the given measurement point for the observation.

c. $\exists J \in$ Envs(x): SingleNogood ⊂ J. The SingleNogood includes in a set from Envs(x). We evaluate this version as $\min_{J \in Envs(x)} (|J| - |SingleNogood|)$ i.e. the less a difference between *SingleNogood* and *Envs(x),* the more priority of a choice of the given measurement point for the current observation.

d. $\forall J \in$ *Envs(x)*: J ∩ SingleNogood = ∅, i.e. no-one of the most probable faulty candidates includes in environments Envs(x) supporting predictions at the point x. We evaluate this version as the least priority choice, i.e. 0 in the numerical equivalent.

Also to the version (d) there are referred other methods of definition of current measurement point priorities which happen when SingleNogood = ∅. Thus in the estimations of a choice priority a numerical value returned as a result of call of other method is accepted. We call it by *ResultD(x).*

At appearance of the greater priority choosing between versions (b) and (c), heuristically we accept the version (b) as at this choice the refinement of faulty candidates is produced better. Note for various supporting sets of the same *Envs*(*x*), the availability both the version (b) and the version (c) is also possible. In this case, as a resulting estimation for the given *Envs*(*x*) the version (b) is also accepted.

Continuing the example, we obtain the following:

Nogood = {{And2,Inv1}, {And2,Inv2,Inv5}, {And2,Inv3}}

SingleNogood = {And2}

< R3, version (c), $\min_{J \in Envs(x)} (|J| - |SingleNogood|) = 2$>

< R1, version (c), $\min_{J \in Envs(x)} (|J| - |SingleNogood|) = 2$>

< O6, version (c), $\min_{J \in Envs(x)} (|J| - |SingleNogood|) = 1$>

< O5, version (a) >

< O4, version (c), $\min_{J \in Envs(x)} (|J| - |SingleNogood|) = 1$>

< O3, version (a) >

< O2, version (c), $\min_{J \in Envs(x)} (|J| - |SingleNogood|) = 1$>

< O1, version (a) >.

Let's estimate the obtained results.
Designate by *maxd* the maximal numerical value among versions (d) for all assessed measurement points, and by *CompCount* a quantity of device components.
Accept in reviewing the following assessments:

1. $\max_{J \in Envs(x)} |J \cap SingleNogood| < CompCount$. The quantity of components which are the intersection result is always less than the quantity of whole device components;
2. $\min_{J \in Envs(x)} (|J| - |SingleNogood|) < CompCount$. The quantity of components in the prediction environment is always less than the quantity of the device components.

Taking into account these assessments, one can introduce a numerical assessment of the obtained results:

$$QuanSNG(x) = \begin{cases} 0, \text{ if } \forall J \in Envs(x) : J \cap SingleNogood = \varnothing \\ ResultD(x), \text{ if } SingleNogood = \varnothing \\ maxD + CompCount - \min\limits_{J \in Envs(x)} (|J| - |SingleNogood|), \\ \quad \text{if } \exists J \in Envs(x) : SingleNogood \subset J \\ maxD + CompCount + \max |J \cap SingleNogood|, \\ \quad \text{if } \exists J \in Envs(x) : |J \cap SingleNogood| < |SingleNogood| \\ maxD + 2 * CompCount, \text{ if } \exists J \in Envs(x) : J \equiv SingleNogood \end{cases}$$

Accordingly, for the example in the Fig. 9:
maxD = 0;
CompCount = 11;
QuanSNG(R3) = 0 + 11 - 2 = 9
QuanSNG(R1) = 0 + 11 - 2 = 9
QuanSNG(O6) = 0 + 11 - 1 = 10
QuanSNG(O5) = 0 + 2*11 = 22
QuanSNG(O4) = 0 + 11 - 1 = 10
QuanSNG(O3) = 0 + 2*11 = 22
QuanSNG(O2) = 0 + 11 - 1 = 10
QuanSNG(O1) = 0 + 2*11 = 22
The points with the greatest value of function *QuanSNG*(*x*) have the greatest priority of choice. We will call the given method as SCAIEH (Supporting and Coinciding Assumptions of Inconsistent Environment Heuristics).
One can see that the most preferable measurement points are O1, O3 and O5, one of environment assumption of which coincides with *SingleNogood*. It differs from guidelines of other choice methods, but at the same time for the given example the choice of any of these points allows to test the most probable faulty component And2.
The developed methods of heuristic choice of the best current measurement point are recommended to use for devices with a great quantity of components as quality of guidelines directly depends on the quantitative difference of environments.

7.5 Practical results

Let's test the developed methods of the best measurement point choosing for the 9-bit parity checker [Frohlich, 1998].

For each experiment one of device components is supposed working incorrectly what is exhibited in a value on its output opposite predicted. A consequence of the incorrect component work is changing of outputs of those components which produce the results depending on values on the output of a faulty component. These changed results of component operations are transmitted to appropriate inquiries of a diagnostic system.
At the beginning of each experiment to inputs of components (Invl, Inv2, Inv3, Inv7, Inv8, Inv9, Invl3, InvI4, Invl5) in a diagnostic complex the vector of values (1,0,1, 0,1,0, 1,0,1) enters. Then to the diagnostic system the value 0 retrieved from the output of the component Nor5 that depends on the work of a broken component and differs from predicted is transferred. It leads to the appearance of an inconsistency in the diagnostic system and starts the automatic process of testing.
In Fig. 10 the quantity of the stages required to each method for fault localization is shown. A method stage is a measurement point choosing. The smaller the quantity of method stages, the faster a fault is localized.

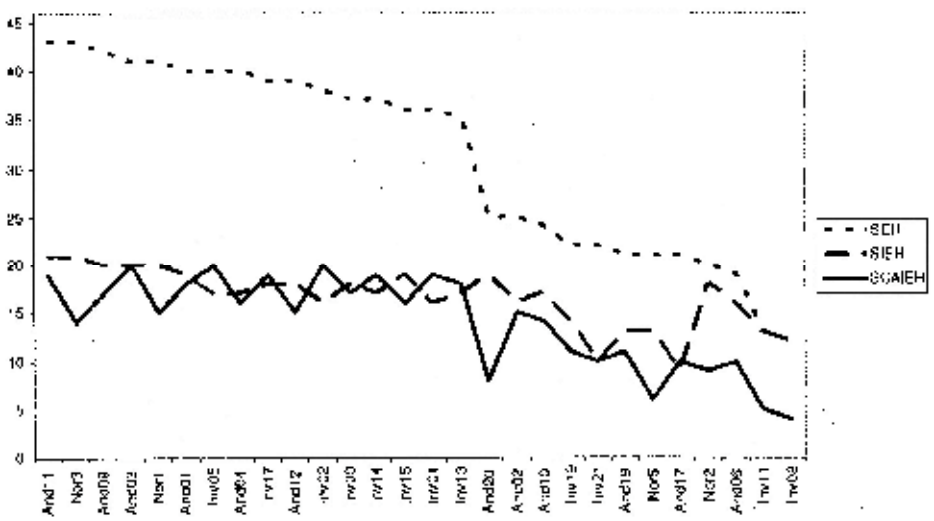


Fig. 10. The quantity of stages required to each method for fault localization

From the obtained results one can see that the method efficiency for different fault components is various and hardly depends on the device structure.
Let's estimate the method efficiency. The device is consists of 46 components. The output values of 45 components are unknown (a value on the output of Nor5 is transmitted to the diagnostic system with input data together). So, the maximal stage quantity necessary for a fault definition is equal 45. Let's accept 45 stages as 100 %. For each experiment it is computed on how many percents each of the developed methods is more effective than exhaustive search of all values. Then define the average value of results. The evaluated results are represented in table 1.

The method	SEH	SEH	SCAIEH
On how many percents the method is more effective, %	30,79	63,17	68,65

Table 1. Results of experiments

From table 1 one can see that the greatest efficiency of current measurement point choosing has the heuristic method based on the knowledge about coincided assumptions of the inconsistent environments SCAIEH.

8. Conclusion

The method of reasoning by analogy on the basis of structural analogy was considered from the aspect of its application in modern IDSS, in particular, for a solution of problems of real-time diagnostics and forecasting. The example of the algorithm for solution search on the basis of analogy of properties that takes into account the context was proposed. This algorithm uses a modified structure of analogy that is capable of taking into account not one property (as in the base algorithm), but a set of properties. These properties determine the original context of analogy and transfer from the source to the receiver only those facts that are relevant in the context of the constructed analogy.

The method of case-based reasoning was considered from the aspect of its application in modern IDSS and RT IDSS, in particular, for a solution of problems of real-time diagnostics and forecasting. The CBR-cycle is considered and its modification for application in RT IDSS is offered. The k-nearest neighbor algorithm for definition of similarity degree of the current situation with cases from case library is described. The structure of case library for RT IDSS is proposed. The proposed method of case-based reasoning was implemented in Borland C++ Builder 6.0 for Windows 7/NT/XP. The main functional components of the implemented tool (Case Libraries Constructor – CLC) are specified.

The possibility of application of analogous reasoning in case-based reasoning is underlined. We also note that the elements of case-based reasoning may be used successfully in analogy-based reasoning methods; i.e., these methods successfully complement each other and their integration in IDSS is very promising.

The heuristic methods of finding the best current measurement point based on environments of device components work predictions are presented.

Practical experiments have confirmed the greatest efficiency of current measurement point choosing for the heuristic method based on the knowledge about coincided assumptions of the inconsistent environments SCAIEH.

Advantages of heuristic methods of the best current measurement point choosing is the simplicity of evaluations and lack of necessity to take into consideration the internal structure interconnections between components of the device.

The presented methods and tools were applied at implementation of a prototype of RT IDSS on the basis of non-classical logics for monitoring and control of complex objects like power units and electronic circuits.

9. References

Vagin, V.N., Eremeev, A.P., 2001. *Some Basic Principles of Design of Intelligent Systems for Supporting Real-Time Decision Making* // Journal of Computer and Systems Sciences International, v. 40(6), p.953-961.

Vagin V.N., Yeremeyev A.P., 2007. *Modeling Human Reasoning in Intelligent Decision Support Systems* // Proc. of the Ninth International Conference on Enterprise Information Systems. Volume AIDSS. Funchal, Madeira, Portugal, June 12-16, INSTICC, pp.277-282.

Haraguchi, M., Arikawa, S., 1986. *A Foundation of reasoning by Analogy. Analogical Union of Logic Programs* // Proceedings of Logic Programming Conference, Tokyo.

Long, D., Garigliano, R., 1994. *Reasoning by analogy and causality: a model and application* // Ellis Horwood Series in Artificial Intelligence.

Varshavskii, P.R., Eremeev, A.P., 2005. *Analogy-Based Search for Solutions in Intelligent Systems of Decision Support* // Journal of Computer and Systems Sciences International, v. 44(1), p. 90–101.

Eremeev, A., Varshavsky, P., 2005. *Analogous Reasoning for Intelligent Decision Support Systems* // Proceedings of the XIth International Conference "Knowledge-Dialogue-Solution" – Varna, v.1, p. 272-279.

Eremeev, A.P. , Varshavsky, P.R., 2006a. *Reasoning by structural analogy in intelligent decision support systems* // Proceedings of the VIIth Joint Conference on Knowledge-Based Software Engineering JCKBSE'06. IOS Press, pp. 303-306.

Aamodt, E., 1994. *Plaza "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches"* // AI Communications, No. 7.

Eremeev, A., Varshavsky, P., 2006b. Analogous Reasoning and Case-Based Reasoning for Intelligent Decision Support Systems // International Journal INFORMATION Theories & Applications (ITHEA), Vol.13 № 4, pp. 316-324.

Eremeev, A., Varshavsky, P., 2007a. *Application of Case-based reasoning for Intelligent Decision Support Systems* // Proceedings of the XIII-th International Conference "Knowledge-Dialogue-Solution" – Varna, vol. 1, pp. 163-169.

Eremeev, A., Varshavsky, P., 2007b. *Methods and Tools for Reasoning by Analogy in Intelligent Decision Support Systems* // Proc. of the International Conference on Dependability of Computer Systems. Szklarska Poreba, Poland, 14-16 June, IEEE, pp.161-168.

Eremeev, A., Varshavsky, P., 2008. *Case-based reasoning method for real-time expert diagnostics systems* // International Journal "Information Theories & Applications", Volume 15, Number 2, pp. 119-125.

Clancey W., 1985. Heuristic Classification // Artificial Intelligence, 25(3), pp. 289-350.

de Kleer, J. and Williams, B.C.,1987. Diagnosing multiple faults // Artificial Intelligence, v.32, pp. 97-130.

Forbus K.D., de Kleer, J. 1993. Building Problem Solver // A Bradford Book, The MIT Press, Cambridge, Massachusetts, London, England.

de Kleer J. , 1986. An Assumption-based TMS // Artificial Intelligence, v.28, p.127-162.

Vagin V.N., Golovina E.Ju., Zagoryanskaya A.A., Fomina M.V., 2008. Exact and Plausible Inference in *Intelligent Systems / Vagin V.N., Pospelov D.A. (eds.)* // M.: Fizmatlit, - 710 p. (in Russian).

Vagin V.N., Oskin P.V., 2006a. The Heuristic Methods of Obtaining the Effective Measurement in Diagnostic Systems // Knowledge-based Software Engineering Proceedings of the 7th Joint Conference on Knowledge-based Software Engineering./ E. Tyugu and T. Yamaguchi (eds). IOS Press, pp.275-284.

Vagin V.N., Os'kin P.V., 2006b. Heuristic and Probabilistic Methods for Taking Efficient Readings in Diagnostic Systems // Journal of Computer and Systems Sciences International. V.l. 45, No 4, pp. 584 -598.

Vagin V.N., Os'kin P.V., 2006c. Multiagent Simulation Subsystem of Diagnostic Complexes Based on Device Models // Journal of Computer and Systems Sciences International. Vol. 45, No 6, pp. 970-982.

Frohlich P., 1998. DRUM-II Efficient Model-based Diagnosis of Technical Systems // PhD thesis, University of Hannover.

Part 2

Hybrid Methodology for System Development

7

Using CBR as Design Methodology for Developing Adaptable Decision Support Systems

Hugo López-Fernández, Florentino Fdez-Riverola,
Miguel Reboiro-Jato, Daniel Glez-Peña and José R. Méndez
University of Vigo
Spain

1. Introduction

Although knowledge-based systems (KBS), and more generally decision support systems (DSS), represent one of the commercial successes resulting from artificial intelligence (AI) research, their developers have repeatedly encountered several problems covering their whole life cycle (Watson, 1997). In this context, knowledge elicitation as well as system implementation, adaptation and maintenance are non trivial issues to be dealt with. With the aim of overcoming these problems, Schank (1982) proposed a revolutionary approach called case-based reasoning (CBR), which is in effect, a model of human reasoning. The idea underlying CBR is that people frequently rely on previous problem-solving experiences when facing up new problems. This assertion may be verified in many day to day problem-solving situations by simple observation or by psychological experimentation (Klein & Whitaker, 1988). Since the ideas underlying case-based reasoning were first established, CBR systems have been found to be successful in a wide range of application areas (Kolodner, 1993; Watson, 1997; Pal et al. 2000). Motivated by the outstanding achievements obtained, some relevant conferences (i.e. ECCBR[1] and ICCBR[2]) and international journals (e.g.: International Journal Transactions on Case-Based Reasoning) have successfully grown up in the field.
In this chapter we present key aspects related with the application of CBR methodology to the construction of adaptable decision support systems. The rest of the chapter is organized as follows: Section 2 introduces an overview about CBR life cycle and combination strategies for constructing hybrid AI systems. Section 3 introduces and covers the main characteristics of four successful decision support systems developed following CBR principles. Finally, Section 4 summarizes the main conclusions and presents the fundamental advantages of adopting this methodology.

2. CBR life cycle and combination strategies

A case-based reasoning system solves new problems by adapting solutions that were used to solve previous problems (Riesbeck & Schank, 1989). The case base holds a number of

[1] http://www.eccbr.org/
[2] http://www.iccbr.org/

cases, each of which represents a problem together with its corresponding solution. Once a new problem arises, a possible solution to it is obtained by retrieving similar cases from the case base and studying their recorded solutions. A CBR system is dynamic in the sense that, in operation, cases representing new problems together with their solutions are added to the case base, redundant cases are eliminated and others are created by combining existing cases.

Every time a CBR system copes with a new problem situation, retrieves previously stored cases together with their solutions, matches them against the new problem context, adapts previous outcomes to provide an answer to the new problem and stores the new solution by adding a novel case in the case base. All of these actions are self-contained and can be represented by a cyclic sequence of processes (see Figure 1) in which human interaction may be needed. A typical CBR system is composed of four sequential steps which are called into action each time a new problem is to be solved (Watson, 1997; Kolodner, 1993; Aamodt & Plaza, 1994): (*i*) *retrieve* the most relevant case(s) (*ii*) *reuse* the case(s) in an attempt to resolve the problem, (*iii*) *revise* the proposed solution if necessary and (*iv*) *retain* the new solution as a part of a new case.

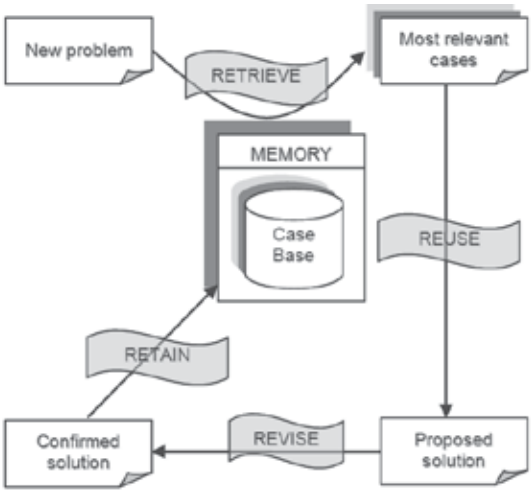


Fig. 1. Typical CBR life cycle comprising four stages

Each of the steps comprising the CBR life cycle defined in Figure 1 requires a model or method in order to accurately perform its mission (see Figure 2). The purpose of the retrieval step is to search the case base and select one or more previous cases that most closely match the new problem situation, together with their solutions. The selected cases are reused to generate a solution appropriate to the current problem situation. This solution is revised if necessary and finally, the new case (i.e. the problem description together with the obtained solution) is stored in the case base. In the CBR cycle there is normally some human interaction. Whilst case retrieval and reuse may be automated, case revision and retention are often undertaken by human experts. This is a current weakness of CBR systems and one of their major challenges. As showed in Figure 2, the techniques commonly used for implementing the different stages of a typical CBR system include: knowledge-based systems, artificial neural networks (ANN), genetic algorithms (GA), rule-based systems (RBS), qualitative reasoning (QR), fuzzy systems (FS) and constraint satisfaction problems (CSP).

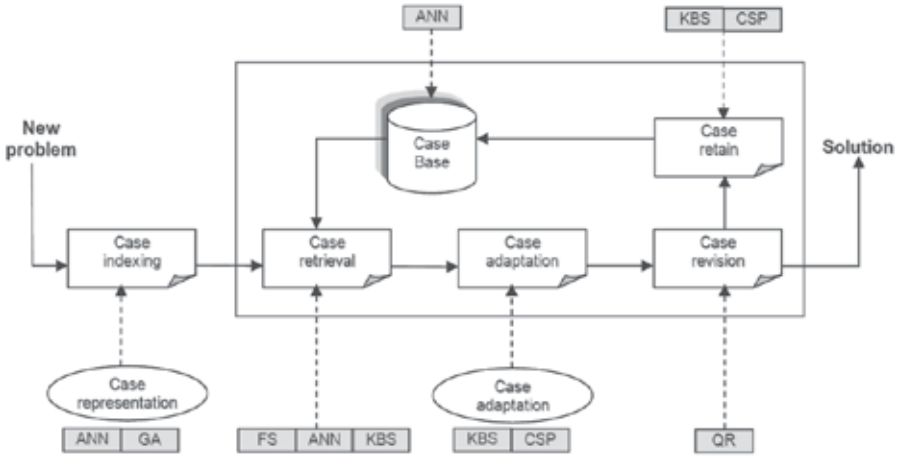


Fig. 2. Methods and techniques commonly used for implementing each stage of a typical CBR system

As the core of CBR systems is its memory, cases should therefore accurately represent both problems and their solutions. In this context, cases may be deleted if they are found to produce inaccurate solutions, they may be merged together to create more generalised solutions, and they may be modified, over time, through the experience gained in producing improved solutions. If an attempt to solve a problem fails and it is possible to identify the reason for the failure, then this information should also be stored in order to avoid the same mistake in the future. This corresponds to a common learning strategy employed in human problem-solving. Rather than creating general relationships between problem descriptors and conclusions, as is the case with rule-based reasoning, or relying on general knowledge of the problem domain, CBR systems are able to utilise the specific knowledge of previously experienced, concrete problem situations. A CBR system provides an incremental learning process because each time a problem is solved, a new experience is retained, thus making it available for future reuse.

As design methodology adequate for constructing DSS, case-based reasoning can be used by itself or as part of another intelligent or conventional computing system. From the last years, there has been an increasing interest in the possibility of integrating different AI techniques with the goal of constructing more powerful and accurate hybrid systems. In this context, the work of Soucek (1991) established the IRIS (*Integration of Reasoning, Informing and Serving*) classification with the goal of facilitating the efficient design of intelligent systems. In the same line, the work of Medsker & Bailey (1992) proposed five integration models based on symbolic and connectionistic systems. Finally, the work of Bezdek (1994) suggested the CIC (*Computational Intelligence Classification*) schema, an interesting classification guidelines for cataloguing hybrid AI systems.

Therefore, given their hybrid nature, CBR systems can be easily combined with other alternatives in order to construct robust decision support systems. These approaches include their successful hybridization with expert systems (Vo & Macchion, 1993; Rissland et al. 1993; Medsker, 1995), fuzzy logic (Xu, 1994; Gui, 1993; Dubois et al. 1997), genetic algorithms (Louis et al. 1993; Oppacher & Deugo, 1991), qualitative reasoning (Navinchandra et al. 1991), constraint satisfaction systems (Maher & Zhang, 1993; Hinrichs, 1992), artificial neural networks (Thrift, 1989; Lim et al. 1991; Liu & Yan, 1997; Corchado et al. 2001) and bayesian

networks (Shinmori, 1998; Aamodt y Langseth, 1998; Dingsøyr, 1998; Friese, 1999; Langseth et al., 1999) among others.

Furthermore, case-based reasoning can be a particularly appropriate problem-solving strategy when the knowledge required to formulate a rule-based model of the domain is difficult to obtain, or when the number or complexity of rules relating to the problem domain is too large for conventional knowledge acquisition methods. In this sense, according to the work of Aamodt and Plaza (1994) there are five different types of CBR systems, and although they share similar features, each of them is more appropriate for a particular type of problem: (*i*) exemplar based reasoning (EBR), (*ii*) instance based reasoning (IBR), (*iii*) memory-based reasoning (MBR), (*iv*) analogy-based reasoning (ABR) and (*v*) typical case-based reasoning (CBR).

EBR systems are especially suitable for classification tasks, in which the category of the most similar past case becomes the solution to the new problem. The set of classes directly constitutes the collection of possible solutions applied without modification. IBR systems are a specialization of exemplar-based reasoning for solving problems in which the instances (cases) are usually very simple (e.g. feature vectors). These systems can be completely automated with no user intervention in their whole life cycle. MBR systems supplement previous approaches with the capacity of parallel processing computation. ABR systems are particularly applicable for solving new problems based on past cases from a different domain. In order to properly work, it should be possible to transfer the solution of a source analogue situation to the present target problem. In typical CBR systems cases are assumed to present a certain complexity level (in terms of their internal structure and the information contained), therefore some modification is needed in order to adapt retrieved solutions when applied to a different problem solving context.

3. Practical applications

The decision support systems covered in this chapter come from four different research areas: industrial planning, biomedical domain, oceanographic forecasting and anti-spam filtering. All the implemented applications are fully designed following the CBR paradigm in order to empower their adaptability and accuracy for solving new problems in their respective fields.

For each domain, we first introduce the target problem to be solved together with the main aspects surrounding each particular situation. A clear description of the representation used for defining the case base is presented and the internal architecture governing each system is explained in detail.

3.1 Industrial planning

Production scheduling is one of the most important functions in a production company. As a consequence, in recent decades various methods have been proposed for the modelling and solution of particular scheduling problems (Akyol & Bayhan, 2007). In the particular case of cooperative poultry farms, the accurate coordination of centralized feed supply (production and distribution) between scattered farms is of utmost importance for both the main feed manufacturer and participating farmers.

In such a situation, some key aspects involving the main participants need to be taken into consideration, for example (*i*) the feed production plant has a limited production and storage capacity, (*ii*) the plant manufactures several types of feed that can create resource

conflicts in production (i.e. they can not be treated as one single product) and distribution (i.e. they can not share the same storage areas in vehicles), (*iii*) deliveries to farmers can be made in advance, but not late, (*iv*) each farm has a limited storage area leading to inventory holding costs, (*v*) some vehicles can not access certain farms due to their dimensions, etc. The problem is difficult since it combines two sub-problems known to be NP-hard: a multi-product, multi-period production problem and a split delivery periodic vehicle routing problem (Boudia, 2008).

In this context, a worthwhile objective for the main feed production plant of any cooperative is to determine, on a daily basis, (*i*) the amount produced for each type of feed, (*ii*) the quantities of feed distributed to each farmer and (*iii*) the associated delivery trips to minimize total cost over the horizon (setup, holding and distribution costs). In this line we present the SP4 (*System for Prediction & Planning of Provision Production*) application, a decision support system that combines a statistical method (used to calculate the previous consumption data, mortality indices and feed delivery types), a machine learning method (M5 Prime and IBk models - used to calculate the total amount of feed consumed by type) and an ad-hoc algorithm which makes flexible orders for compound feed production forecasting (Reboiro-Jato et al. 2011).

Attribute	Type	Description
Farm Id	Alphanumeric	Farm identifier.
Shed Id	Numeric	Shed identifier. Each farm may have several sheds. This value differentiates each shed in a given farm.
Entry date	dd-mm-yyyy	Date on which the lot arrives at the farm.
Final date	dd-mm-yyyy	Date on which the lot is slaughtered.
Feed type	Enumeration	Identifies the specific variety of feed.
Breed	Enumeration	Specific breed type of the lot.
Activity	Enumeration	This variable indentifies the main role of the lot. There are five possible values: *laying hen raising, breeder raising, laying hen laying, breeder laying and fattening*.
Subactivity	Enumeration	Activity specialization. Each activity may have several specific variants. Possible values include: *broiler, label, light, country*, etc.
Type	Enumeration	Class of fatten chicken. Animals used on fatten activity are classified into several categories like 'A', 'C', 'CA', 'H', 'M', etc. This attribute is only used on lots belonging to *fatten* activity.
Number of males	Numeric	Number of male animals the lot.
Number of females	Numeric	Number of female animals the lot.
Number of unsexed	Numeric	Number of animals for which the sex is unknown.
Season	Enumeration	Period of the year in which the lot spends the most days on the farm.
Kilograms	Numeric	Amount of feed consumed by the lot from its entry to its slaughter.

Table 1. Variables defining the SP4 internal case base

Efficient farming of animal husbandry on such an industrial scale relies largely on several variables and limitations that are normally correlated. In such a situation, characterized by multiple scattered farms that depend on a centralized supply centre, there are two main subjects to take into account: (*i*) different breeds of poultry growing in the same farm (many animals) with a distinct raise program and (*ii*) different types of feed needed for each lot in each farm for a given period of time. In order to build an appropriate knowledge base for the accurate operation of the SP4 system, several economic, production and management company databases were studied and pre-processed. The main objective was gathering and filtering real data in order to obtain a valid history of feed orders sent to farmers during recent years. Table 1 presents the final structure of the cases comprising the memory of the SP4 system. The general criterion applied for the development of the SP4 internal case base was to select those orders sent to the farmer in the period comprising four days before each lot arrived at the farm, until its exit. During this process, several rules were codified and applied for detecting and solving existing inconsistencies.

Given the large number of variables and rules that play an important role in the whole process of correctly supplying feed for all the farms comprising a given cooperative, the proposed decision support system has been organized into three different but complementary subsystems: (*i*) data gathering and feed type identification, (*ii*) consumption forecasting and (*iii*) production planning. The developed system is going to execute subsystems one and two in a sequential fashion for each new lot of animals that arrives to a given farm. Finally, subsystem three is responsible for combining all this information on a daily basis and determining the amount of feed that must be produced by day and type to the main feed production plant. Figure 3 depicts the general architecture of the SP4 system.

As it can be seen from the operational diagram showed in Figure 3, each subsystem works with unrelated information coming from both existing data and knowledge bases, but also considering the output of the previous stage. In the proposed architecture, each phase solves a different problem following a bottom-up approach.

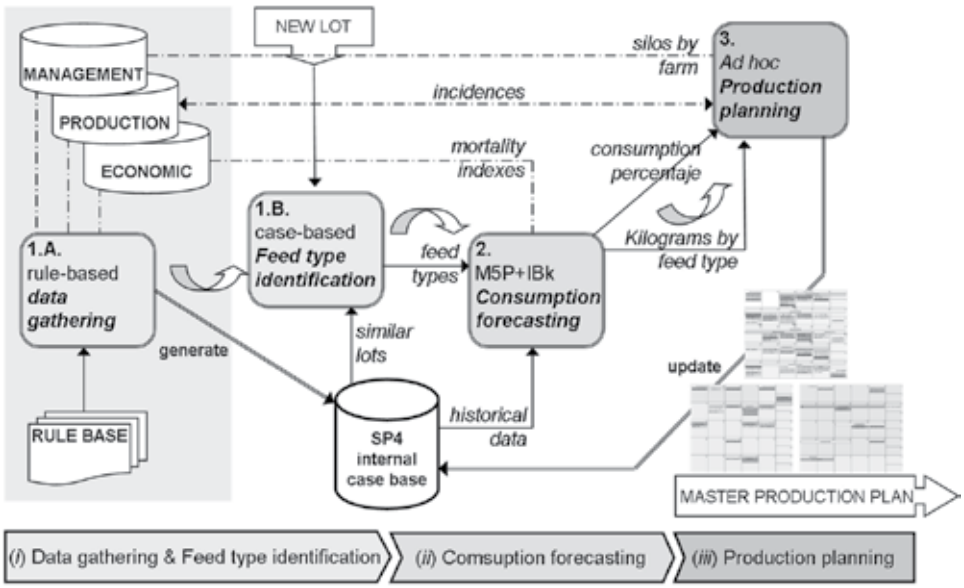


Fig. 3. General architecture and life cycle of the SP4 system for compound feed forecasting and planning

Subsystem one: each lot of animals consumes several types of feed during each growth stage. The objective of this phase is to identify the variants of feed that a given lot is going to consume before slaughter, taking into consideration similar past cases belonging to different farms and periods. The current output (qualitative information) serves as the basis for the next phase to compute the total amount of feed needed by both a given lot and a given feed type (quantitative information).

Subsystem two: given a new lot and the different types of feed that it will consume before slaughter, it estimates the total consumption (kilograms) by feed type together with the consumption percentage per animal. Several variables influence the output of this phase: number of animals by sex (male and female), week of year, mortality indexes, etc.

Subsystem three: this integrates previous calculations by using an ad hoc planning algorithm for simulating feed consumption and production along with the time. It allows the user to easily visualize production activity though an intuitive and configurable GUI (*Graphic User Interface*). Incidences in the initial estimated MPS can be reported and accounted for in reconsidering daily production.

The data used for constructed the system case base was provided by a leading Spanish company specialized in animal feed production and delivery. Raw data (from the years 2007 and 2008) was built from client orders, company production logs, information about the number of animals at different farms and truck trips to the clients. A total of 5112 records were stored in the SP4 internal case base after executing the initial operation for detecting and solving existing inconsistencies (procedure 1.A. in Figure 3). In order to increase the confidence of experimental findings, a cross-validation process was carried out in the all of the experiments.

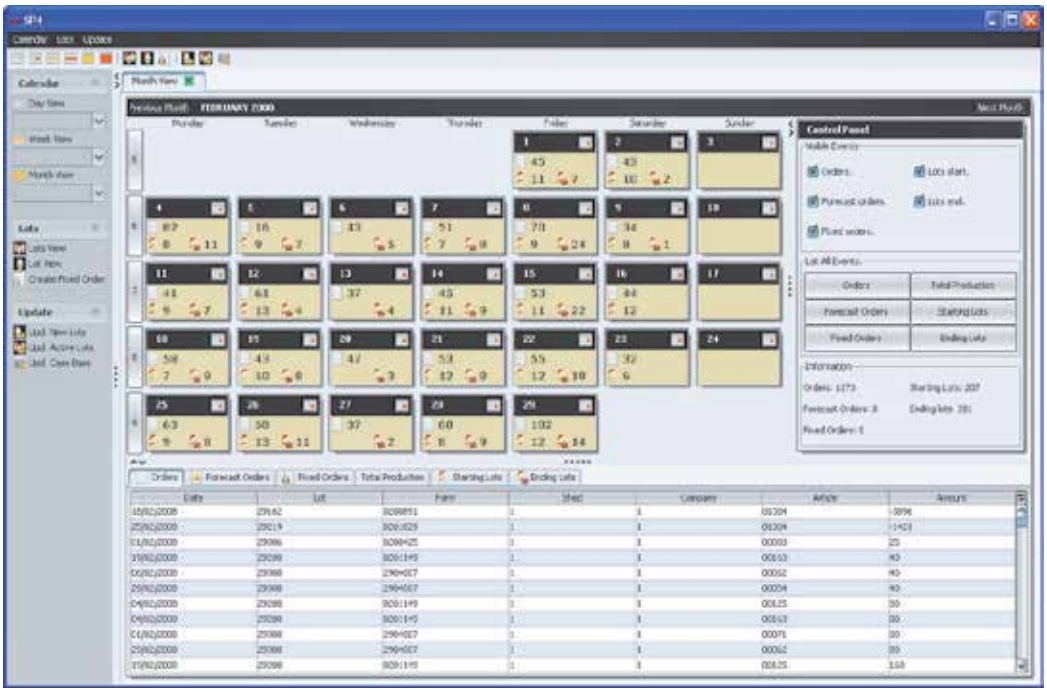


Fig. 4. Screenshot of the final deployed SP4 system showing the Master Production Plan

The final developed system used in the experiments carried out (see Figure 4) was set up to forecast the total amount (kilograms) of each feed type consumed by each lot from its entry date to its slaughter. Moreover, further experiments have been carried out to compare the performance of the SP4 system with several other forecasting approaches. These include standard statistical forecasting algorithms, decision trees and the application of neural networks methods. The results obtained from experimentation revealed that the proposed system performed optimally, being able to track the dynamic non-linear trend and seasonality, as well as the numerous interactions between correlated variables.

3.2 Biomedical domain

In recent years DNA microarray technology has become a fundamental tool in genomic research, making it possible to investigate global gene expression in all aspects of human disease (Russo et al. 2003). Microarray technology is based on a database of over 40,000 fragments of genes called expressed sequence tags (ESTs), which are used to measure target abundance using the scanned intensities of fluorescence from tagged molecules hybridized to ESTs. Following the advent of high-throughput microarray technology it is now possible to simultaneously monitor the expression levels of thousands of genes during important biological processes and across collections of related samples. Since the number of examined genes in an experiment runs to the thousands, different data mining techniques have been intensively used to analyze and discover knowledge from gene expression data (Piatetsky-Shapiro & Tamayo 2003). However, having so many fields relative to so few samples creates a high likelihood of finding false positives. This problem is increased if we consider the potential errors that can be present in microarray data.

Bioinformatics and medical informatics are two research fields that serve the needs of different but related communities. Both domains share the common goal of providing new algorithms, methods and technological solutions to biomedical research, and contributing to the treatment and cure of diseases. Although different microarray techniques have been successfully used to investigate useful information for cancer diagnosis at the gene expression level, the true integration of existing methods into day-to-day clinical practice is still a long way off (Sittig et al. 2008). Within this context, case-based reasoning emerges as a suitable paradigm specially intended for the development of biomedical informatics applications and decision support systems, given the support and collaboration involved in such a translational development (Jurisica & Glasgow, 2004).

In addressing the issue of bridging the existing gap between biomedical researchers and clinicians who work in the domain of cancer diagnosis, prognosis and treatment using microarray data, we have developed and made accessible a common interactive framework: the geneCBR decision support system (Glez-Peña et al. 2009a). Our geneCBR system implements a freely available software tool that allows the use of combined techniques that can be applied to gene selection, clustering, knowledge extraction and prediction for aiding diagnosis in cancer research. For biomedical researches, geneCBR *expert mode* offers a core workbench for designing and testing new techniques and experiments. For pathologists or oncologists, geneCBR *diagnostic mode* implements an effective and reliable system that can diagnose cancer subtypes based on the analysis of microarray data using CBR architecture. For programmers, geneCBR *programming mode* includes an advanced edition module for run-time modification of previous coded techniques.

In order to initially construct the knowledge base starting from the available patient's data showed in Table 2, geneCBR stores the gene expression levels of each microarray sample in its case base (lower part of Figure 5).

Attribute	Type	Description
NAME	Alphanumeric	Unique identifier. Denomination of the microarray sample.
#Age	Numeric	Age of the patient.
#Sex	Enumeration	Possible values are: *male*, *female*.
#FAB/WHO	Alphanumeric	The FAB classification is a morphological characterization of leukemia. The WHO classification incorporates the results of chromosome and genetic research developed during the past 25 years after the FAB classification.
#FISH studies	Alphanumeric	FISH studies are used to delineate complex chromosome rearrangements, diagnose microdeletion syndromes, or demonstrate the presence of molecular rearrangements characteristic of certain hematologic malignancies.
Gene name *1*	Alphanumeric	Human gene name or identifier.
Gene value *1*	Numeric	Microarray gene expression value for the associated gene identifier.
….	…	The number of gene name-value pairs depends on the microarray type (Affymetrix HG-U133A/B/Plus, etc.).
Gene name *n*	Alphanumeric	Human gene name or identifier.
Gene value *n*	Numeric	Microarray gene expression value for the associated gene identifier.
Class	Alphanumeric	Type of disease.

Table 2. Internal representation of a microarray sample in the geneCBR system (the symbol '#' represents an optional feature)

During the retrieval stage, the original case vectors are transformed into fuzzy microarray descriptors (FMDs). Each FMD is a comprehensible descriptor of the sample in terms of a linguistic label for each gene expression level (central part of Figure 5). This transformation is carried out by means of an accurate fuzzy discretization process (Díaz et al. 2006). Based on the FMD representation created from the case base, a set of fuzzy patterns (FP) is constructed that represents the main characteristics of the a priori known classes (top-left square in Figure 5). Each class in the system is then represented by a FP that holds the fuzzy codification of gene expression levels for those genes that were flagged as relevant for this class. Several FPs are generated from the data in a supervised way, each one representing a group of FMDs for pathological (or generic) specy.

The retrieval stage in the geneCBR system uses the FPs in order to select the most representative genes given a new patient. This phase can be thought of as a gene selection step, in which the aim is to retrieve the list of genes that might be most informative given a new sample to classify. Since it is highly unlikely that all the genes have significant information related to cancer classification and the dimensionality would be too great if all the genes were used, it is necessary to explore an efficient way to obtain the most suitable group of genes. In order to make this selection, our geneCBR system selects those fuzzy patterns from its case base which are the nearest to any new case obtained. Then, for each one of the selected FPs, the geneCBR system computes its associated DFP (a pattern which only includes the genes that are necessary in order to discriminate the novel instance from other different classes). Finally, the selected genes for the new case are obtained by joining together the genes belonging to the DFPs considered.

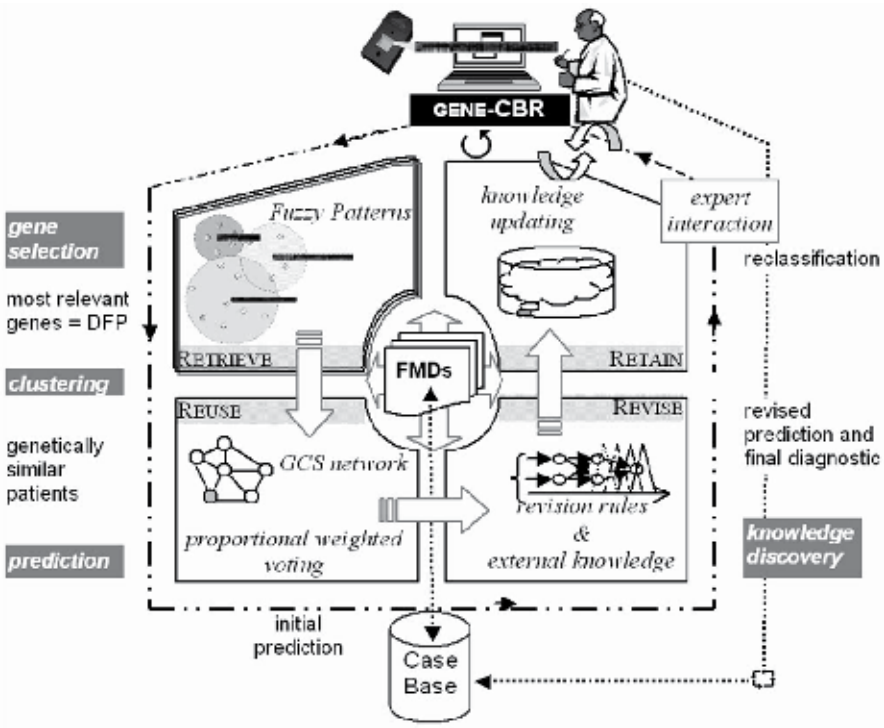


Fig. 5. Life cycle of the geneCBR system working in *diagnostic mode*

The adaptation of previous cases in order to solve a new FMD is accomplished in the reuse stage (left bottom square in Figure 5). A growing cell structures (GCS) network (Fritzke, 1993) is trained with the whole case base, only taking the existing cases represented by the genes selected in the previous stage as input. Then, the new FMD is presented to the GCS network and the patients most similar from a genetic point of view are retrieved. Based on this grouping, a proportional weighted voting mechanism is applied that ponders the level of similarity with the new FMD. An initial class is assigned by the geneCBR system from among the existing pathologies (Glez-Peña et al. 2009b).

In the revision stage (right bottom square in Figure 5) the expert is provided with useful data about the decision made by the system. This information contains the selected DFP genes, the grouping made by the GCS network, the weighting assigned to each class and a set of See5 (Quinlan, 2000) classification rules generated from the most similar patients. The expert contrasts the initial prediction given by the system with other external information such as patient karyotype or clinical history in order to ascertain a revised prediction and a final diagnostic.

Every time a new FMD is solved, the internal structure of the geneCBR system is updated (top-right square in Figure 5). The new FMD is associated to its corresponding class and added to the case base. The affected FP is updated and the system marks the most similar patients selected for future classifications. In this stage the geneCBR system changes to edit-mode and the expert is permitted to update patient's classification taking into account the new knowledge generated.

The geneCBR represents a new translational decision support system that can effectively support the integrative work of programmers, biomedical researches and clinicians working

together in a common framework. Figure 6 shows a screenshot of the geneCBR system working in ex*pert mode* (specially intended for biomedical researches). The code of the project is freely available under the GPL license and can be obtained at http://www.genecbr.org/.

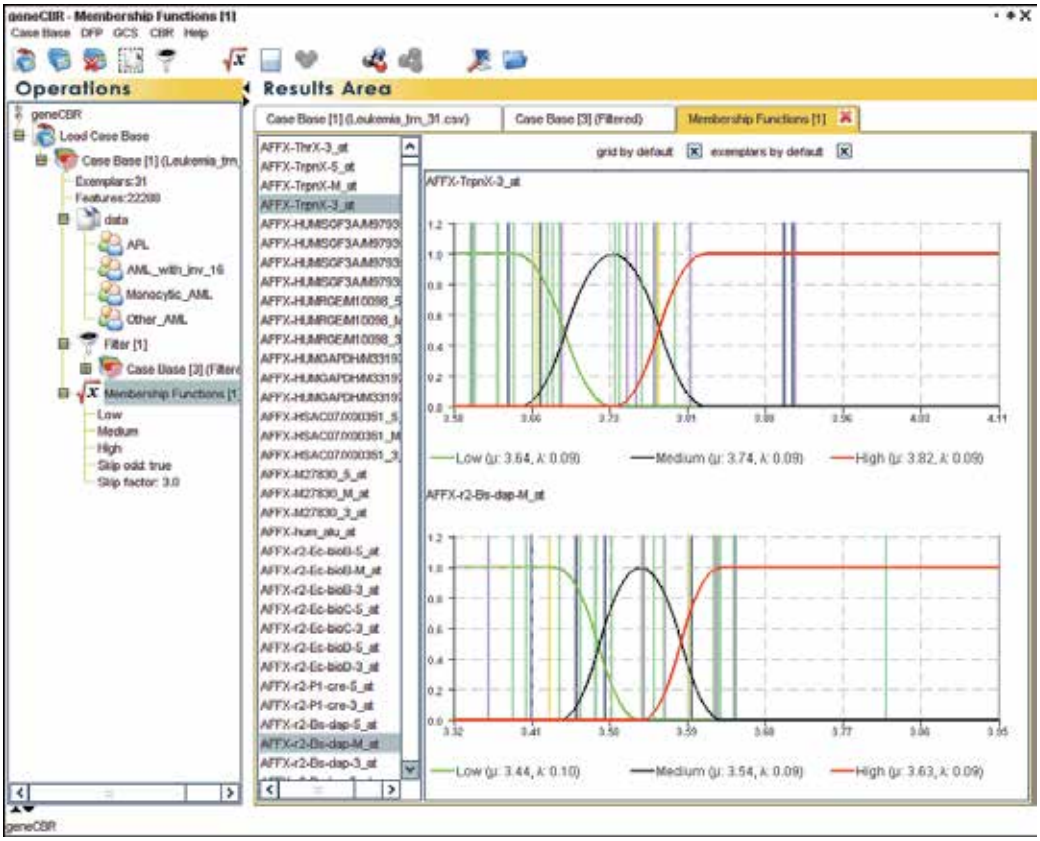


Fig. 6. Screenshot of the geneCBR system working in *expert mode*

3.3 Oceanographic forecasting

The oceans of the world form a highly dynamic system for which it is difficult to create mathematical models (Tomczak & Fodfrey, 1994). *Red tides* are the name for the discolourations caused by dense concentrations of microscopic sea plants, known as phytoplankton. The discolouration varies with the species of phytoplankton, its pigments, size and concentration, the time of day, the angle of the sun and other factors. Red tides usually occur along the North West coast of the Iberian Peninsula in late summer and autumn. The prevailing southerly winds cause cold, nutrient-rich water to rise up from the deeper regions of the ocean to the surface, a process known as upwelling. Swept along with this upwelled water are dinoflagellate cysts, the resting stages of the organism, which lie dormant in the sediments on the sea floor. The high nutrient concentrations in the upwelled water, together with ideal conditions of temperature, salinity and light, trigger the germination of the cysts, so that the dinoflagellates begin to grow and divide. The rapid increase in dinoflagellate numbers, sometimes to millions of cells per liter of water, is described as a bloom of phytoplankton (concentration levels above the 100,000 cells/liter).

Concentration of the bloom by wind and currents, as well as the dinoflagellates' ability to swim to the surface, combines to form a red tide.

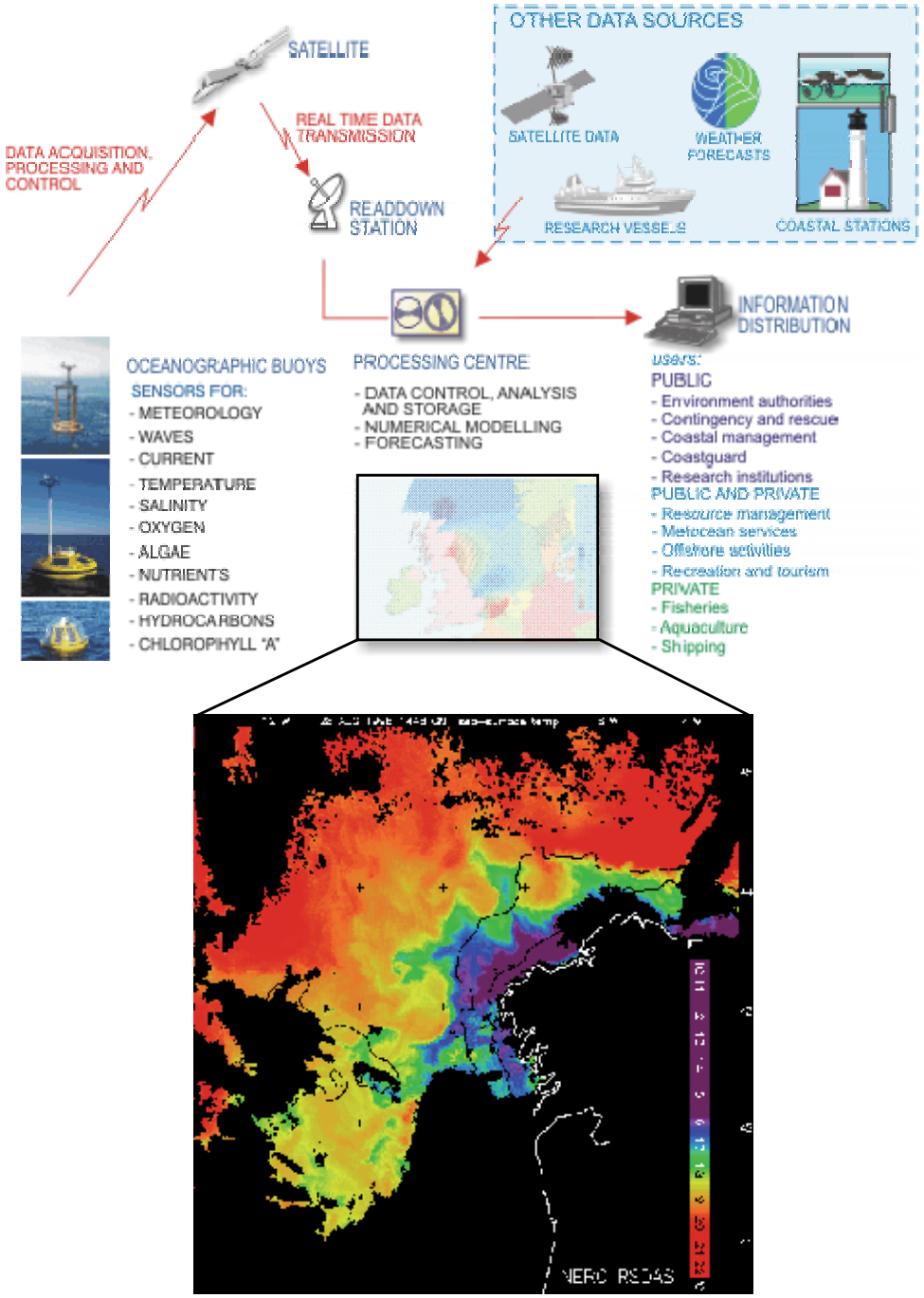


Fig. 7. Data sources used by the FSfRT sytem

In this context, two situations of special interest are those corresponding to the *false alarms* and the *blooms not detected*. The former refers to predictions of bloom (concentration of pseudo-nitzschia ≥ 100,000 cell/liter) which do not actually materialize (real concentration ≤ 100,000 cell/liter). The latter, more problematic, occurs when a bloom exists but the model fails to detect it. Another unwelcome situation takes place when the number of predictions exceeds an absolute error of 100,000 cell/liter (labelled as incorrect predictions). In such a situation, in which the rules governing a target process or system are unknown, the prediction of the parameter values that determine the characteristic behaviour of the system can be a problematic task. However, it has been found that a hybrid case-based reasoning system can provide a more effective means of performing such predictions than other connectionist or symbolic techniques (Fdez-Riverola & Corchado, 2003).

Related with this domain we describe the FSfRT (*Forecasting System for Red Tides*) system, a hybrid model able to accurately forecast the concentrations of pseudo-nitzschia spp, the diatom that produces the most harmful red tides causing amnesic shellfish poisoning (or ASP). Our FSfRT system employs a case-based reasoning model to wrap a growing cell structures network, a radial basis function network (Fritzke, 1994) and a set of Sugeno fuzzy models (Jang et al. 1997) to provide an accurate prediction. Each of these techniques is used at a different stage of the reasoning cycle of the decision support system to retrieve historical data, to adapt it to the present problem and to automatically review the proposed solution.

The forecasting system uses information from two main sources: (*i*) data coming from several buoys and monitoring net used to create a succession of problem descriptors able to characterize the current forecasting situation and (*ii*) data derived from satellite images stored on a database. The satellite image data values are used to generate cloud and superficial temperature indices which are then stored with the problem descriptor and subsequently updated during the CBR operation. Figure 7 shows a schematic view of the whole data managed by the FSfRT system.

In order to forecast the concentration of pseudo-nitzschia spp at a given point a week in advance, a problem descriptor is generated on a weekly basis. A problem descriptor consists of a sequence of sampled data values (filtered and pre-processed) recorded from the water mass to which the forecast will be applied. The problem descriptor also contains various other numerical values, including the current geographical location of the sensor buoys and the collection time and date. Every week, the concentration of pseudo-nitzschia spp is added to a problem descriptor forming a new input vector. The problem descriptor is composed of a vector with the variables that characterise the problem recorded over two weeks. The prediction or output of the system is the concentration of pseudo-nitzschia spp one week later, as indicated in Table 3.

The cycle of forecasting operations (which is repeated every week) proceeds as depicted in Figure 8. First a new problem instance is created from the pre-processed data cited above. When a new problem is presented to the system, the GCS neuronal network is used to obtain *k* more similar cases to the given problem (identifying the class to which the problem belongs). In the reuse phase, the values of the weights and centers of the neural network used in the previous forecast are retrieved from the knowledge base. These network parameters together with the *k* retrieved cases are then used to retrain the RBF network and to obtain an initial forecast of the concentration of pseudo-nitzschia spp (see Figure 8). During this process the values of the parameters that characterise the network are updated.

Attribute	Type	Description
Location	Alphanumeric	Geographical location of the sensor buoy.
Date	dd-mm-yyyy	Date on which the measure was made.
Time	hh-mm-ss	Time on which the measure was made.
Temperature	Numeric	Water temperature (cent. degrees) at different depths.
Oxygen	Numeric	Oxygen concentration (milliliters/liter) at different depths.
PH	Numeric	acid/based scale.
Transmitance	Numeric	Fraction (percentage) of sun light that passes through the sea-water.
Fluorescence	Numeric	Sea-water fluorescence (percentage) at different depths.
Cloud index	Numeric	Cloud measurement derivate from a geostationary satellite.
Recount of diatoms	Numeric	Algae concentration (in cell/liter) at different depths.
Pseudo-nitzschia spp	Numeric	Diatom concentration (in cell/liter) at different depths causing harmful algae blooms.
Pseudo-nitzschia spp (future)	Numeric	Diatom concentration (in cell/liter) to be predicted.

Table 3. Oceanographic parameters and physical characteristics of the water mass comprising a case in the FSfRT system

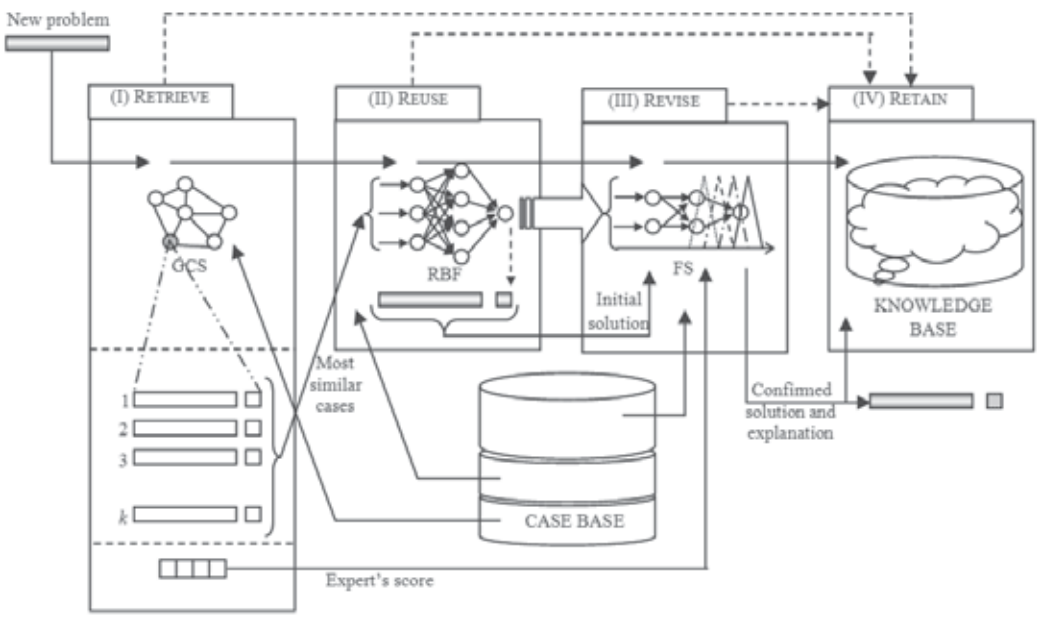


Fig. 8. Life cycle of the FSfRT system for oceanographic prediction

In the revision phase, the initial solution proposed by the RBF neural network is modified according to the responses of the four fuzzy revision subsystems. Each revision subsystem has been created from the RBF network using neurofuzzy techniques (Jin & Sendhoff, 2003). For each class of the GCS neural network a vector of four values is maintained (see Figure

8). This *expert's score* vector is initialised with a value of (0.25, 0.25, 0.25, 0.25) and represents the accuracy of each revision subsystem with respect to a class. During revision, the appropriate *expert's score* vector is used to ponder the outputs of each fuzzy revision system. Each vector value is associated with one of the four revision subsystems. For each forecasting cycle, the value of the importance vector associated to the most accurate revision subsystem is increased and the other three values are proportionally decreased. This is done in order to give more relevance to the most accurate revision subsystem.

The revised forecast is then retained temporarily in the forecast database. When the real value of the concentration of pseudo nitzschia spp is measured, the forecast value for the variable can then be evaluated, though comparison of the actual and forecast value and the error obtained. A new case, corresponding to this forecasting operation, is then stored in the case base. The forecasting error value is also used to update the importance vector associated with the revision subsystems of the retrieved class.

The FSfRT system was successfully tested using real data collected from years [1992, 2000] coming from geographical area A0 (42º28.90′ N, 8º57.80′ W 61 m). Figure 9 shows a screenshot of the FSfRT interface implemented for oceanographic forecasting.

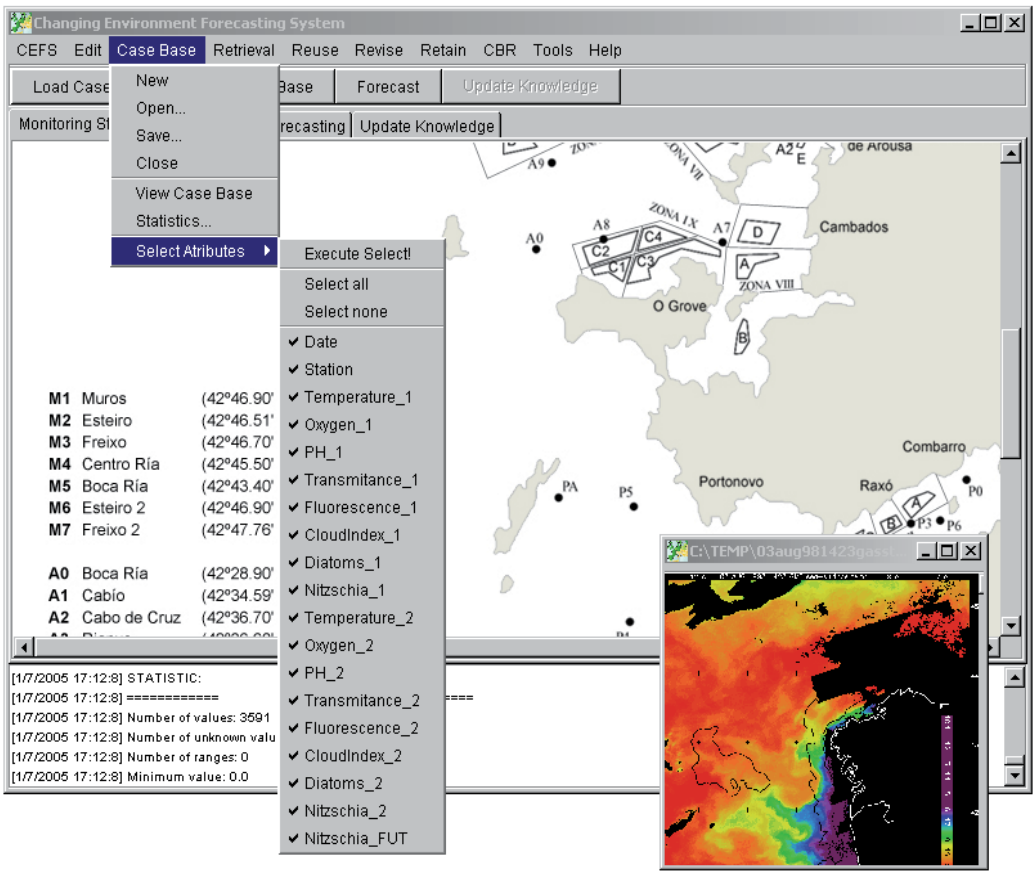


Fig. 9. Screenshot of the FSfRT system forecasting red tides in the coastal waters of the North West of the Iberian Peninsula

3.4 Anti-spam filtering

The E-mail service is a computer-based technology built as the result of transforming the old postal delivery in order to use it over networks and Internet. Nowadays, e-mail addresses are present on every business card close to other relevant contact info such as the postal address or the phone number. However, for more than one decade the use of e-mail has been bedeviled by the curse of spamming, so spam is beginning to undermine the integrity of e-mail and even to discourage its use.

In this context, spam is a term used to designate all forms of unsolicited commercial e-mail and can be formally defined as an electronic message satisfying the following two conditions: (*i*) the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients and (*ii*) the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent (SpamHaus, 1998).

Due to some attractive characteristics of e-mail (low cost & fast delivery) it actually becomes the main distribution channel of spam contents. Every day e-mail users receive lots of messages containing offers to buy illegal drugs, replicas of Swiss watches, fake jobs, forged university diplomas, etc. This situation has led to a progressive increasing of the spam global ratio in email traffic. During September 2010, the percentage of spam deliveries accounted for about 92 percent of all Internet e-mail traffic (MessageLabs, 2010).

In order to successfully fight against spam (i.e. ideally eliminate it), both theoretical and applied research on spam filtering becomes fundamental. In this context, much valuable research work has been previously carried out (Guzella & Caminhas, 2009) and some relevant conferences have grown up in the field (CEAS, 2010). Moreover, several commercial products have been released and distributed from the software industry to a huge amount of final users with the goal of minimizing spam drawbacks.

With the goal of providing an effective solution we present the SpamHunting system (Fdez-Riverola et al. 2007), an instance-based reasoning e-mail filtering model that outperforms classical machine learning techniques and other successful lazy learner's approaches in the domain of anti-spam filtering. The architecture of the decision support filter is based on a tuneable enhanced instance retrieval network able to accurately generalize e-mail representations. The reuse of similar messages is carried out by a simple unanimous voting mechanism to determine whether the target case is spam or not. Previous to the final response of the system, the revision stage is only performed when the assigned class is spam whereby the system employs general knowledge in the form of meta-rules.

In order to correctly represent incoming e-mails, a message descriptor (instance) is generated and stored in the e-mail base of the SpamHunting system. This message descriptor contains the sequence of features that better summarize the information contained in the e-mail. For this purpose, we use data from two main sources: (*i*) information obtained from the header of the e-mail and (*ii*) those terms that are more representative of the subject, body and attachments of the message. Table 4 summarizes the structure of each instance stored in the SpamHunting e-mail base.

Figure 10 illustrates the life cycle of the IBR SpamHunting system as well as its integration within a typical user environment. In the upper part of Figure 10, the mail user agent (MUA) and the mail transfer agent (MTA) are in charge of dispatching the requests generated by the user. Between these two applications, SpamHunting captures all the incoming messages (using POP3 protocol) in order to identify, tag and filter spam.

Attribute	Type	Description
ID	Numeric	Unique message identifier.
From	Alphanumeric	Source mailbox.
Return Path	Alphanumeric	Indicates the address that the message will be returned to if one chose to reply.
Date	dd-mm-yyyy	Date in which the message was sent.
Language	Alphanumeric	Particular tongue of the message.
Attached Files	Numeric	Indicates the number of attached files.
Content Type	Enumeration	MIME type.
Relevant Terms	Numeric	Number of selected features to cluster the message.
Total Terms	Numeric	Number of features contained in the message.
Frequency-Term Descriptor	Array of feature-frequency pairs	Storing for each feature a measure of their frequency in the message.
Class	Enumeration	Message category. Possible values are: *spam*, *legitimate*, *unknown*.

Table 4. Structure of an instance representing an incoming e-mail in the SpamHunting system

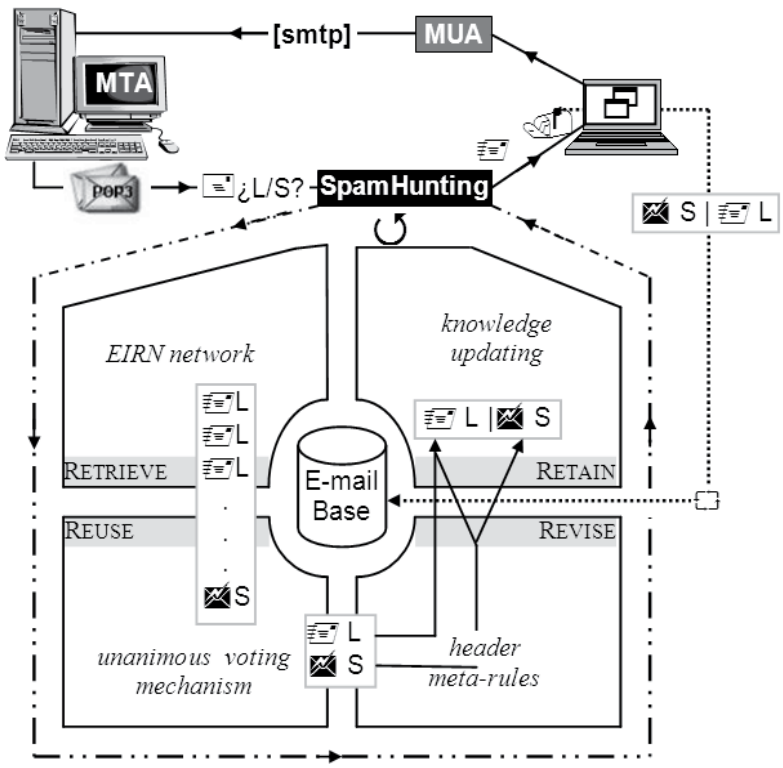


Fig. 10. Life cycle of the SpamHunting system and its integration with existing MTA and MUA

Whenever SpamHunting receives a new e-mail, the system evolves through the four steps depicted in the lower part of Figure 10 as shadowed rectangles. Initially the system

identifies those e-mails that best represent the new incoming message (left upper quadrant in Figure 10), only taking into account the set of messages with the highest number of terms in common. Each one of the previous selected messages contributes with one vote to the final class (left bottom quadrant in Figure 10). The revision stage is only carried out when the system proposes an e-mail as spam. For this purpose, SpamHunting uses previous encoded header meta-rules (right bottom quadrant in Figure 10). Every time the user checks his mailbox and provides feedback to the system about a previous e-mail classification, SpamHunting stores a new instance (or modifies an existing one) in the e-mail base for future use (right upper quadrant in Figure 10).

The retrieval stage is carried out using our enhanced instance retrieval network (EIRN) model. The EIRN network facilitates the indexation of instances and the selection of those that are most similar to the instance-message. The reuse of similar e-mails is carried out by means of the utilization of a unanimous voting mechanism, which generates an initial solution by creating a model with the retrieved instances. In the revision stage the system employs general knowledge in the form of meta-rules that are extracted from the e-mail headers. Finally, the retain (learning) stage is carried out whenever the system classifies an incoming e-mail, updating the knowledge structure of the whole system. The proposed system also takes into account the feedback of the user when it receives an incorrectly classified e-mail (dotted arrow from the user computer to the e-mail base in Figure 10).

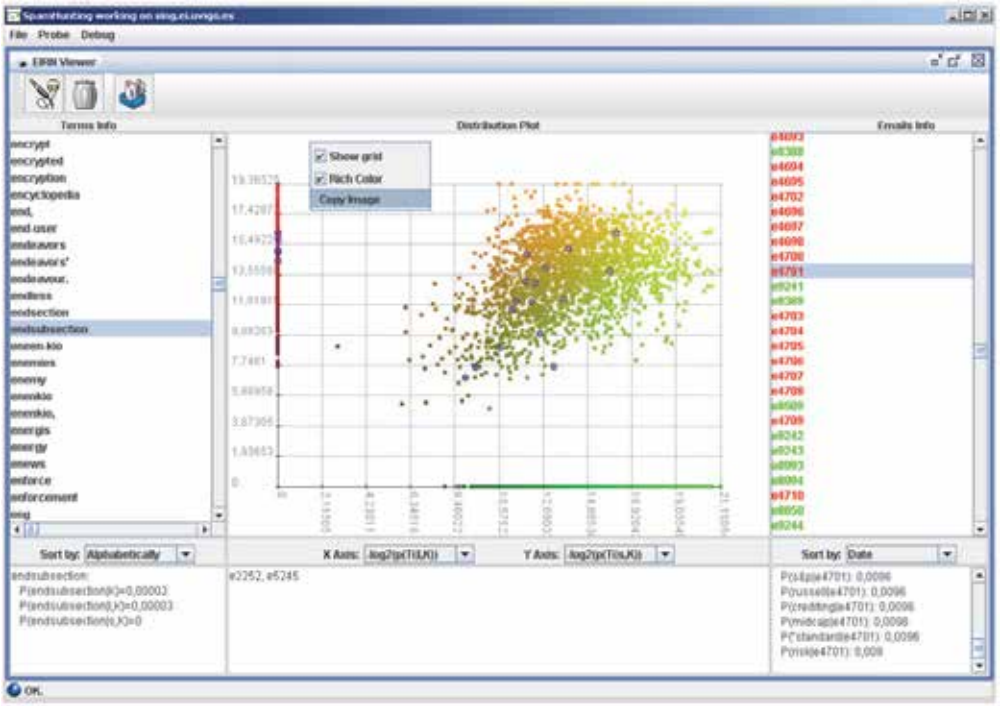


Fig. 11. Screenshot of the EIRN viewer comprising the SpamHunting system

One of the most relevant features of our EIRN structure is the possibility of generating different views of the underlying indexed knowledge. Figure 11 shows a screenshot of the software developed for taking advantage of this capacity. As we can observe from Figure 11,

the application frame is structured into six panels. First, on the top of the left column, we can find all the relevant terms that the EIRN network uses for indexing existing instances. Under this panel, a section is located that summarizes some statistics referring to the selected terms including: (*i*) probability of finding the term in the case base (*ii*) frequency of finding the term in spam e-mails and (*iii*) probability of finding the term in legitimate messages.
At the top of the right column, we can find all the instances stored in the system memory. Under this panel, an instance statistics panel can be found where all the relevant terms belonging to the selected messages as well as their frequencies are showed.
At the top of the central column, we have placed a plot representing the relevant terms indexed in our EIRN model. The plot panel has been developed with built-in clipboard copy support, save images capability and drag and drop compatibility for use with any image application or word processor. Both selected terms belonging to the left panel or part of a message from the right panel are always highlighted in our graphic representation.
The EIRN viewer module represents each term as a two-dimension point on the plot. The coordinates of each point are computed according to the function selected in the combo boxes placed under the plot. The following measurements are available for each coordinate: (*i*) the probability of finding the term t in spam e-mails stored in the system memory, $p(t \mid s, K)$, (*ii*) the logarithmic form of the previous value, $-\log_2(p(t \mid s, K))$ (*iii*) the probability of finding the term t in legitimate messages, $p(t \mid l, K)$, (*iv*) the logarithmic form of the previous value, $-\log_2(p(t \mid l, K))$, and (*v*) the probability of finding the term t in the system memory, $p(t \mid K)$.
The results obtained confirm the idea that instance-based reasoning systems can offer a number of advantages in the spam filtering domain. Spam is a disjoint concept and IBR classification works well in this domain. In addition IBR systems can learn over time simply by updating their memory with new instances of spam or legitimate e-mail. Moreover, it provides seamless learning capabilities without the need for a separate learning process and facilitates extending the learning process over different levels of learning. The code can be freely obtained at http://www.spamhunting.org/.

4. Conclusion

In this chapter we have presented the CBR paradigm as an appropriate methodology for implementing successful decision support systems together with their adequate application to several domains. The adoption of CBR methodology for constructing decision support systems has several remarkable benefits, allowing us to obtain a more general knowledge of the system and to gain a deeper insight into the logical structure of the problem and its solution. Main properties of these systems include (*i*) their ability to focus on the problem's essential features, (*ii*) the possibility of solving problems in domains that are only partially understood, (*iii*) the competence for providing solutions when no algorithmic method is available and (*iv*) the potential to interpret and manage open-ended and ill-defined concepts.
The main advantages of case-based reasoning paradigm over other alternative approaches for the implementation of decision support systems are related with the fulfilment of the following characteristics: (*i*) reduce the knowledge acquisition effort (cases are independent from each other), (*ii*) require less maintenance effort (partially automatically by adding/deleting cases), (*iii*) improve problem solving performance through reuse, (*iv*) makes use of existing data (e.g. in databases), (*v*) improve over time and adapt to changes in the environment and (*vi*) present high user acceptance (domain experts and novices understand cases quite easy).

However, the effective utilization of case-based reasoning paradigm also presents inherent limitations and/or drawbacks related to its knowledge representation and life cycle: (*i*) past cases could be inexistent or difficult to represent, (*ii*) specific techniques have to be defined from scratch for modifying previous cases or their solutions in order to adapt them to the new situations and (*iii*) in some scenarios, it could be difficult to maintain case-base efficiency because unused cases need to be forgotten. In such situations, specific solutions have to be defined in order to overcome particular difficulties.

5. Acknowledgements

This work has been partially funded by the project InNoCBR (10TIC305014PR) from Xunta de Galicia. M. Reboiro-Jato was supported by a pre-doctoral fellowship from University of Vigo.

6. References

Aamodt, A. & Langseth, H. (1998). *Integrating Bayesian Networks into Knowledge-Intensive CBR*. Technical Report WS-98-15, AAAI Press, Menlo Park, pp. (1-6)

Aamodt, A. & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications,* Vol.7, pp. (39–59)

Akyol, D.E. & Bayhan, G.M. (2007). A review on evolution of production scheduling with neural networks. *Computers & Industrial Engineering*, Vol.53, No.1, pp. (95-122)

Bezdek, J.C. (1994). What is Computational Intelligence?. In: *Computational Intelligence: Imitating Life*, Zurada J.M., Marks II, R.J., & Robinson C.J. (Eds.), pp. (1-12), IEEE Press, ISBN 978-0780311046, New York

Boudia, M. (2008). Coordination of production planning and distribution. *4OR: A Quarterly Journal of Operations Research*, Vol.6, No.1, pp. (93-96)

CEAS, (2010). *Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, Available from http://www.ceas.cc/

Corchado, J.M., Aiken, J. & Rees, N. (2001). *Artificial Intelligence Models for Oceanographic Forecasting*. Plymouth Marine Laboratory, ISBN 0-951-9618-4-5, UK

Díaz, F., Fdez-Riverola, F. & Corchado, J.M. (2006). geneCBR: a Case-Based Reasoning Tool for Cancer Diagnosis using Microarray Datasets. *Computational Intelligence*, Vol.22, No.3-4, pp. (254-268)

Dubois, D., Esteva, F., García, P., Godo, L., López de Mántaras, R. & Prade, H. (1997). Fuzzy set-based models in case-based reasoning. *Proceedings of the 2nd International Conference on Case-Based Reasoning*, Providence, Rhode Island, USA, pp. (599-610)

Dingsøyr, T. (1998). *Retrieval of Cases by using a Bayesian Network*. Technical Report WS-98-15, AAAI Press, Menlo Park, pp. (50-54)

Fdez-Riverola, F. & Corchado, J.M. (2003). CBR based System for Forecasting Red Tides. *International Journal of Knowledge-Based Systems*, Vol.16, No.5-6, pp. (321-328)

Fdez-Riverola, F., Iglesias, E.L., Díaz, F., Méndez, J.R. & Corchado, J.M. (2007). SpamHunting: An Instance-Based Reasoning System for Spam Labelling and Filtering. *Decision Support Systems*, Vol.43, No.3, pp. (722-736)

Friese, T. (1999). Utilization of Bayesian Belief Networks for Explanation-Driven Case-Based Reasoning. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, Workshop ML-5: On the automatic construction of case-based reasoners, Stockholm, Schweden, pp. (73-76)

Fritzke, B. (1993). Growing self-organising Networks – Why?, *Proceedings of the 11th European Symposium on Artificial Neural Networks*, Bruges, Belgium, pp. (61-72)

Fritzke, B. (1994). Fast learning with incremental RBF Networks. *Neural Processing Letters*, Vol.1, No.1, pp. (2–5)

Glez-Peña, D., Díaz, F., Hernández, J.M., Corchado, J.M. & Fdez-Riverola, F. (2009a). geneCBR: a translational tool for multiple-microarray analysis and integrative information retrieval for aiding diagnosis in cancer research. *BMC Bioinformatics*. 10:187

Glez-Peña, D., Díaz, F., Fdez-Riverola, F., Méndez, J.R. & Corchado J.M. (2009b). Fuzzy Patterns and GCS Networks to Clustering Gene Expression Data. In: *Fuzzy Systems in Bioinformatics and Computational Biology. Studies in Fuzziness and Soft Computing*, pp. (103-125), Springer, ISBN: 978-3-540-89967-9, Germany

Gui, J.K. (1993). Methodology for modelling complete product assemblies. Act Polytechnica Scandinavica. *Mathematics and Computer Science*, Vol.60 pp (1-189)

Guzella, T.S. & Caminhas, W.M. (2009) A review of machine learning approaches to Spam filtering. *Expert Systems with Applications*, Vol.36, No.7, pp. (10206-10222)

Hinrichs, T.R. (1992). *Problem solving in open worlds: A case study in design*. Lawrence Erlbaum, ISBN 0805812288, Hillsdale, New York

Jang, J.-S.R., Sun, C.-T. & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, ISBN 978-0132610667, US

Jin, Y. & Sendhoff, B. (2003). Extracting Interpretable Fuzzy Rules from RBF Neural Networks. *Neural Processing Letters*, Vol.17, No.2, pp. (149-164)

Jurisica, I. & Glasgow, J. (2004). Applications of case-based reasoning in molecular biology. *Artificial Intelligence Magazine, Special issue on Bioinformatics*, Vol.25, No.1, pp. (85–95)

Klein, G. & Whitaker, L. (1988). Using analogues to predict and plan. *Proceedings of the first DARPA Workshop on Workshop Case-Based Reasoning*, pp. (224–232)

Kolodner, J. (1993). *Case-based Reasoning*. Morgan Kaufmann, ISBN 1-55860-237-2, San Mateo, CA

Langseth, H., Asmodt, A. & Martin, O. (1999). Learning Retrieval Knowledge from Data. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, Workshop ML-5: On the automatic construction of case-based reasoners, Stockholm, Schweden, pp. (77-82)

Lim, J.H., Tan, A.H., Lui, H.C. & Teh, H.H. (1991). INCIDE: A connectionist case-based diagnostic expert system that learns incrementally. *Proceedings of the International Joint Conference on Neural Networks*, Singapore, pp. (1693-1698)

Liu, Z.Q. & Yan, F. (1997). Fuzzy neural network in case-based diagnostic system. *IEEE Transactions on Fuzzy Systems*, Vol.5, No.2, pp. (209-222)

Louis, S., McGraw, G. & Wyckoff, R.O. (1993). Case Based Reasoning assisted explanation of genetic algorithm results. *Journal of Experimental and Theoretical Artificial Intelligence*, Vol.5, No.1, pp. (21-37)

Maher, M.L. & Zhang, D.M. (1993). CADSYN: A case based design process model. *Artificial Intelligence for Engineering Design, Analysis, and Manufacture*, Vol.7, No.2, pp. (97-110)

Medsker, L.R. (1995). Hybrid Intelligent Systems. Kluwer Academic Publishers, ISBN 9780792395881

Medsker, L.R., & Bailey, D.L. (1992). Models and guidelines for integrating expert systems and neural networks. In: *Hybrid Architectures for Intelligent Systems*. Kandel, A., &

Langholz, G. (Eds.), pp. (153-171), CRC Press, ISBN 0-8493-4229-5, Boca Raton, Florida

MessageLabs. (2010). *MessageLabs Intelligence*, Available from http://www.messagelabs.co.uk/intelligence.aspx

Navinchandra, D., Sycara, K.P. & Narasimhan, S. (1991). A transformational approach to case-based synthesis. *Artificial Intelligence in Engineering, Manufacturing and Design*, Vol.5, No.1, pp. (31-45)

Oppacher, F. & Deugo, D. (1991). Integrated Case-based Reasoning with genetic algorithms. *Proceedings of the International Symposium on Computational Intelligence*, Milan, Italy, pp. (103-114)

Pal, S.K., Dilon, T.S. & Yeung, D.S. (2000). Soft Computing in Case Based Reasoning. *International Journal of Intelligent Systems*, Vol.16, No.4

Piatetsky-Shapiro, G. & Tamayo, P. (2003). Microarray data mining: facing the challenges. *ACM SIGKDD Explorations Newsletter*, Vol.5, No.2, pp. (1–5)

Quinlan, J.R. (2000). *Data mining tools See5 and C5.0*, Available from: http://www.rulequest.com/see5-info.html/

Reboiro-Jato, M., Glez-Dopazo, J., Glez, D., Laza, R., Gálvez, J.F., Pavón, R., Glez-Peña, D. & Fdez-Riverola, F. (2011). Using Inductive Learning to Assess Compound Feed Production in Cooperative Poultry Farms. *Expert Systems With Applications*, Submitted for Publication

Riesbeck, C.K. & Schank, R.C. (1989). Inside Case-based Reasoning. Lawrence Elrlbaum Ass, ISBN 0-89859-767-6, Hillsdale

Rissland, E.L., Daniels, J.J., Rubinstein, Z.B. & Skalak, D.B. (1993). Case-based diagnostic analysis in a blackboard architecture. *Proceedings of the Eleventh National Conference on Artificial Intelligence*, Washington, DC, pp. (66-72)

Russo, G., Zegar, C. & Giordano, A. (2003). Advantages and limitations of microarray technology in human cancer. *Oncogene*, Vol.22, No.42, pp. (6497-507)

Schank, R.C. (1982). *Dynamic Memory*. Cambridge University Press, ISBN 0-521-24856-2, Cambridge, UK

Shinmori, A. (1998). *A Proposal to Combine Probabilistic Reasoning with Case-Based Retrieval for Software Troubleshooting*. Technical Report WS-98-15, AAAI Press, Menlo Park, pp. (149-154)

Sittig, D.F., Wright, A., Osheroff, J.A., Middleton, B., Teich, J.M., Ash, J.S., Campbell, E. & Bates, D.W. (2008). Grand challenges in clinical decision support. *Journal of Biomedical Informatics*, Vol.41, No.2, pp. (387-92)

Soucek, B. & IRIS group. (1991). *Neural and Intelligent Systems Integration - Fifth and Sixth Generation Integrated Reasoning Information Systems*. John Wiley & Sons, ISBN 978-0-471-53676-5, New York

SpamHaus. (1998). *The SpamHaus Project*, Available from http://www.spamhaus.org/

Thrift, P. (1989). A neural network model for case-based reasoning. *Proceedings of the Second DARPA Workshop on Case-Based Reasoning*, Pensacola Beach, Florida, pp. (334-337)

Tomczak, M. & Godfrey, J.S. (1994). *Regional Oceanographic: An Introduction*. Pergamon, ISBN 978-0080410203, New York

Vo, D.P. & Macchion, D. (1993). A use of case-based reasoning technique in building expert systems. *Future Generation Computer systems*, Vol.9, No.4, pp. (311-319)

Watson, I. (1997). *Applying Case-based Reasoning: Techniques for Enterprise Systems*. Morgan Kaufmann, ISBN 978-1558604629, San Mateo, CA

Xu, L.D. (1994). Developing a case-based knowledge system for AIDS prevention. *Expert Systems*, Vol.11, No.4, pp. (237-244)

8

Identification of Key Drivers of Net Promoter Score Using a Statistical Classification Model

Daniel R. Jeske[1], Terrance P. Callanan[2] and Li Guo[3]
[1]*Professor and Chair, Department of Statistics, University of California, Riverside, CA*
[2]*Customer Experience Manager, Six Sigma Champion, Carestream Health, Rochester, NY*
[3]*Graduate Student, Department of Statistics, University of California, Riverside, CA*
USA

1. Introduction

Net Promoter Score (NPS) is a popular metric used in a variety of industries for measuring customer advocacy. Introduced by Reichheld (2003), NPS measures the likelihood that an existing customer will recommend a company to another prospective customer. NPS is derived from a single question that may be included as part of a larger customer survey. The single question asks the customer to use a scale of 0 to 10 to rate their willingness and intention to recommend the company to another person. Ratings of 9 and 10 are used to characterize so-called 'promoters,' ratings of 0 through 6 characterize 'detractors,' and ratings of 7 and 8 characterize 'passives.' The NPS is calculated as the percentage of respondents that are promoters minus the percentage of respondents that are detractors.

The idea behind the labels given to customers is as follows. Promoters are thought to be extremely satisfied customers that see little to no room for improvement, and consequently would offer persuasive recommendations that could lead to new revenue. The passive ratings, on the other hand, begin to hint at room for improvement and consequently the effectiveness of a recommendation from a Passive may be muted by explicit or implied caveats. Ratings at the low end are thought to be associated with negative experiences that might cloud a recommendation and likely scare off prospective new customers. Additional discussion on the long history of NPS can be found in Hayes (2008).

Some implementations of NPS methodology use reduced 5-point or 7-point scales that align with traditional Likert scales. However it is implemented, the hope is that movements in NPS are positively correlated with revenue growth for the company. While Reichheld's research presented some evidence of that, other findings are not as corroborative (Kenningham et al., 2007). Regardless of whether there is a predictive relationship between NPS and revenue growth, implementing policies and programs within a company that improve NPS is an intuitively sensible thing to do [see, for example, Vavra (1997)]. A difficult and important question, however, is how to identify key drivers of NPS. *Calculating* NPS alone does not do this.

This chapter is an illustrative tutorial that demonstrates how a statistical classification model can be used to identify key drivers of NPS. Our premise is that the classification model, the data it operates on, and the analyses it provides could usefully form components of a Decision Support System that can not only provide both snapshot and longitudinal analyses of NPS performance, but also enable analyses that can help suggest company initiatives aimed toward lifting the NPS.

We assume that the NPS question was asked as part of larger survey that also probed customer satisfaction levels with respect to various dimensions of the company's services. We develop a predictive classification model for customer advocacy (promoter, passive or detractor) as a function of these service dimensions. A novelty associated with our classification model is the optional use of constraints on the parameter estimates to enforce a monotonic property. We provide a detailed explanation of how to fit the model using the SAS software package and show how the fitted model can be used to develop company policies that have promise for improving the NPS. Our primary objective is to teach an interested practitioner how to use customer survey data together with a statistical classifier to identify key drivers of NPS. We present a case study that is based on a real-life data collection and analysis project to illustrate the step-by-step process of building the linkage between customer satisfaction data and NPS.

2. Logistic regression

In this section we provide a brief review of logistic and multinomial regression. Allen and Rao (2000) is a good reference that contains more detail than we provide, and additionally has example applications pertaining to customer satisfaction modeling.

2.1 Binomial logistic regression

The binomial logistic regression model assumes that the response variable is binary (0/1). This could be the case, for example, if a customer is simply asked the question "Would you recommend us to a friend?" Let $\{Y_i\}_{i=1}^n$ denote the responses from n customers, assigning a "1" for Yes and "0" for No. Suppose a number of other data items (covariates) are polled from the customer on the same survey instrument. These items might measure the satisfaction of the customer across a wide variety of service dimensions and might be measured on a traditional Likert scale. We let $\underset{\sim}{x}_i$ denote the vector of covariates for the i-th sampled customer and note that it reflects the use of dummy variable coding for covariates that are categorical scale. For example, if the first covariate is measured on a 5-point Likert scale, its value is encoded into $\underset{\sim}{x}_i$ by using five dummy variables $\{x_{1j}\}_{j=1}^5$, where $x_{1j} = 1$ if and only if the Likert response is j.

The binomial logistic regression model posits that Y_i is a Bernoulli random variable (equivalently, a binomial random variable with trial size equal to one) with success probability p_i, and further, that the success probability is tied to the covariates through the so-called link function $p_i = \exp(\alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i) / [1 + \exp(\alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i)]$, where $\underset{\sim}{\beta}$ is a vector of model parameters (slopes). Continuing with the 5-point Likert scale example above, there would be five slopes $\{\beta_{1j}\}_{j=1}^5$ associated with the five dummy variables $\{x_{1j}\}_{j=1}^5$ used to code the first covariate.

Model fitting for the binomial logistic regression model entails estimating the parameters α and $\underset{\sim}{\beta}$ via maximum likelihood. The likelihood function for this model is

$$L(\alpha,\underset{\sim}{\beta}) = \prod_{i=1}^{n} \frac{\exp[(\alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i)Y_i]}{1+\exp(\alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i)} \tag{1}$$

and the maximum likelihood estimate (MLE) of $(\alpha,\underset{\sim}{\beta}')$ is the value, say $(\hat{\alpha},\hat{\underset{\sim}{\beta}}')$, that maximizes this function. Once the MLE is available, the influence of the covariates can be assessed by the magnitude (relative to their standard error) of the individual slopes. In particular, it can be ascertained which attributes of customer service have a substantial effect on making the probability of a 'Yes' response high.

2.2 Multinomial logistic regression

Suppose now that the outcome variable has more than two categories. For example, suppose the responses $\{Y_i\}_{i=1}^{n}$ are measured on the 11-point NPS scale. The multinomial logistic model seeks to represent the probability of *each* response as a function of the covariates. Since each response is an ordinal categorical variable taking on values in the set $\{0,1,\ldots,10\}$, we consider a multinomial logistic regression model with 10 cumulative link functions:

$$\Pr(Y_i \le j) = \frac{\exp(\alpha_j + \underset{\sim}{\beta}'\underset{\sim}{x}_i)}{1+\exp(\alpha_j + \underset{\sim}{\beta}'\underset{\sim}{x}_i)} \quad , \quad j = 0,1,\ldots,9 \tag{2}$$

where $\{\alpha_j\}_{j=0}^{9}$ are intercept parameters and $\underset{\sim}{\beta}$ is again a vector of slope parameters. In order to affect the required non-decreasing behavior (relative to *j*) of the right hand sides of the link functions, the constraint $\alpha_0 \le \alpha_2 \cdots \le \alpha_9$ is imposed on the intercept parameters. Starting with $\Pr(Y_i = 0) = \Pr(Y_i \le 0)$, and then differencing the above expressions in (2) as in $\Pr(Y_i = j) = \Pr(Y_i \le j) - \Pr(Y_i \le j-1)$, $j = 1,\ldots,10$, we obtain expressions for the individual probabilities of the response as a function of the intercept and slope parameters. Defining the responses $\{Y_{ij}\}_{j=0}^{10}$ to be one (zero) if and only if $Y_i = j$, the likelihood function for this model is

$$L(\underset{\sim}{\alpha},\underset{\sim}{\beta}) = \prod_{i=1}^{n}\prod_{j=0}^{10} P(Y_i = j)^{Y_{ij}} \tag{3}$$

and the MLE of $(\underset{\sim}{\alpha},\underset{\sim}{\beta}')$ is the value, say $(\hat{\underset{\sim}{\alpha}},\hat{\underset{\sim}{\beta}}')$, that maximizes this function. Once the MLE is available, the magnitudes of the slope estimates (relative to their standard errors) can be used to identify the covariates that push the distribution of the response towards 9s and 10s. We note that the constraint on the intercepts is a standard constraint. In the next section, we will discuss an additional and novel constraint that can optionally be imposed on the slope parameters.

3. Case study

Carestream Health, Inc. (CSH) was formed in 2007 when Onex Corporation of Toronto, Canada purchased Eastman Kodak Company's Health Group and renamed the business as Carestream Health. They are an annual $2.5B company and a world leader in medical imaging (digital and film), healthcare information systems, dental imaging and dental practice management software, molecular imaging and non-destructive testing. Their

customers include medical and dental doctors and staff and healthcare IT professionals in small offices and clinics to large hospitals and regional and national healthcare programs. A major company initiative is to create a sustainable competitive advantage by delivering the absolute best customer experience in the industry. Customer recommendations are key to growth in the digital medical space and no one has been able to do it consistently well. The foundation for taking advantage of this opportunity is to understand what's important to customers, measure their satisfaction and likelihood to recommend based on their experiences, and drive improvement.

While descriptive statistics such as trend charts, bar charts, averages and listings of customer verbatim comments are helpful in identifying improvement opportunities to improve the Net Promoter Score (NPS), they are limited in their power. First, they lack quantitative measurements of correlation between elements of event satisfaction and NPS. As a consequence, it is not clear what impact a given process improvement will have on a customer's likelihood to recommend. Second, they lack the ability to view multi-dimensional relationships – they are limited to single factor inferences which may not sufficiently describe the complex relationships between elements of a customer's experience and their likelihood to recommend.

This section summarizes the use of multinomial logistic regression analyses that were applied to 5056 independent customer experience surveys from Jan 2009 – Jan 2010. Each survey included a question that measured (on a 5-point Likert scale) how likely it would be for the customer to recommend colleagues to purchase imaging solutions from CSH. Five other questions measured the satisfaction level (on a 7-point Likert scale) of the customer with CSH services obtained in response to an equipment or software problem. Key NPS drivers are revealed through the multinomial logistic regression analyses, and improvement scenarios for specific geographic and business combinations are mapped out. The ability to develop a quantitative model to measure the impact on NPS of potential process improvements significantly enhances the value of the survey data.

3.1 CSH Customer survey data

The 5-point Likert response to the question about willingness to recommend is summarized in Table 1 below. CSH calculates a unique net promoter score from responses on this variable using the formula $\text{NPS} = \sum_{i=1}^{5} w_i \hat{p}_i$, where $\underset{\sim}{w} = (-1.25, -0.875, -0.25, 0.75, 1.0)'$ is a vector of weights and where $\hat{p}_i$ is the estimated proportion of customers whose recommendation score is *i*. Two interesting characteristics of the weight vector are, first, the penalty for a 1 (2) exceeds the benefit of a 5 (4), and second, the negative weight for a neutral score is meant to drive policies toward delighting customers.

Recommendation	Interpretation
1	Without being asked, I will advise others NOT to purchase from you
2	Only if asked, I will advise others NOT to purchase from you
3	I am neutral
4	Only if asked, I will recommend others TO purchase from you
5	Without being asked, I will recommend others TO purchase from you

Table 1. Meaning of Each Level of Recommendation Score

Let $Y \in \{1,2,3,4,5\}$ be a random variable denoting the willingness of a particular customer to recommend CSH. The values $\{p_i\}_{i=1}^{5}$ represent the theoretical probabilities of the possible values for Y. That is, $p_i = \Pr(Y = i)$. The multinomial logistic regression model treats the customer demographic variables and the customer satisfaction ratings as the covariates $\underset{\sim}{x}_i$, linking their values to the probability distribution of Y such that their influence on the values $\{p_i\}_{i=1}^{5}$ can be ascertained. Since the expected value of NPS is a function of these $\{p_i\}_{i=1}^{5}$ values, the end result is a model that links NPS to what customers perceive to be important. This linkage can then be exploited to determine targeted pathways to improve NPS via improvement plans that are customer-driven.

The demographic covariates include the (global) region code, country code, business code and the customer job title. The demographic covariates are coded using the standard dummy variable coding technique. For example, region code utilizes 7 binary variables $\{RC_i\}_{i=1}^{7}$ where

$$RC_i = \begin{cases} 1 & \text{if customer falls in i-th global region} \\ 0 & \text{otherwise .} \end{cases}$$

Country code utilizes similar dummy variables, but because countries are nested within regions we use the notation $\{CC_{j(i)}\}_{i=1}^{7}{}_{j=1}^{n_i}$, where n_i is the number of country codes within the i-th region and where

$$CC_{j(i)} = \begin{cases} 1 & \text{if customer falls in j-th country within } i\text{-th region} \\ 0 & \text{otherwise .} \end{cases}$$

In the data set we have $\{n_i\}_{i=1}^{7} = \{3,1,1,7,5,4,2\}$. Business code utilizes two dummy variables $\{BC_i\}_{i=1}^{2}$ where

$$BC_i = \begin{cases} 1 & \text{if customer is aligned with } i\text{-th business code} \\ 0 & \text{otherwise .} \end{cases}$$

Finally, job title utilizes 10 dummy variables $\{JT_i\}_{i=1}^{10}$ where

$$JT_i = \begin{cases} 1 & \text{if customer has } i\text{-th job title} \\ 0 & \text{otherwise .} \end{cases}$$

The customer satisfaction covariates are also coded using the dummy variable scheme. The data on these covariates are the responses to the survey questions identified as q79, q82a, q82b, q82d and q82f. These questions survey the customer satisfaction on 'Overall satisfaction with the service event,' 'Satisfaction with CSH knowledge of customer business and operations,' 'Satisfaction with meeting customer service response time requirements,' 'Satisfaction with overall service communications,' and 'Satisfaction with skills of CSH employees,' respectively.

Survey questions q82c and q82e, which survey satisfaction with 'Time it took to resolve the problem once work was started,' and 'Attitude of CSH employees' were also considered as covariates, but they did not show themselves to be statistically significant in the model. Their absence from the model does not necessarily imply they are not important drivers of

their overall satisfaction with CSH, but more likely that their influence is correlated with the other dimensions of overall satisfaction that are in the model. Each customer satisfaction covariate is scored by customers using a 7-point Likert scale (where '1' indicates the customer is "extremely dissatisfied" and '7' indicates "extremely satisfied"), and thus each utilizes 7 dummy variables in the coding scheme. We denote these dummy variables as $\{q79_i\}_{i=1}^7$, $\{q82\mathrm{a}_i\}_{i=1}^7$, $\{q82\mathrm{b}_i\}_{i=1}^7$, $\{q82\mathrm{d}_i\}_{i=1}^7$, and $\{q82\mathrm{f}_i\}_{i=1}^7$, respectively, and they are defined as follows:

$$q79_i = \begin{cases} 1 & \text{if customer response to q79 is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82\mathrm{a}_i = \begin{cases} 1 & \text{if customer response to q82a is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82\mathrm{b}_i = \begin{cases} 1 & \text{if customer response to q92b is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82\mathrm{d}_i = \begin{cases} 1 & \text{if customer response to q82d is } i \\ 0 & \text{otherwise}, \end{cases}$$

$$q82\mathrm{f}_i = \begin{cases} 1 & \text{if customer response to q82f is } i \\ 0 & \text{otherwise}. \end{cases}$$

Assembling all of the covariates together, we then have a total of 77 covariates in $\underset{\sim}{x}$. Thus, the vector of slopes $\underset{\sim}{\beta}$ in the link equations has dimension 77×1. Combined with the 4 intercept parameters $\{\alpha_i\}_{i=1}^4$, the model we have developed has a total of 81 parameters. We note it is conceivable that interactions between the defined covariates could be important contributors to the model. However, interaction effects based on the current data set were difficult to assess because of confounding issues. As the data set gets larger over time, it is conceivable the confounding issues could be resolved and interaction effects could be tested for statistical significance.

3.2 Model fitting and interpretation

The SAS code for obtaining maximum likelihood estimates (MLEs) for the model parameters $\{\alpha_i\}_{i=1}^4$ and $\underset{\sim}{\beta}$ is shown in Appendix A. Lines 1-4 are used to read in the data that is stored as a space delimited text file 'indata.txt' that is located in the indicated directory. All of the input variables on the file are coded as integer values. The PROC LOGISTIC section of the code (lines 5-10) directs the fitting of the multinomial logistic regression model. The class statement is used to specify that all of the covariate variables are categorical in nature, and the param=glm option specifies to use the dummy variable coding scheme that was defined in the previous section. Table 2 summarizes the portion of the SAS output that reports the maximum likelihood estimates for $\{\alpha_i\}_{i=1}^4$ and $\underset{\sim}{\beta}$. Note that the zero for the slope of the last level of each covariate is a structural zero resulting from the non-full rank dummy variable coding used when fitting the model.

Parm.	Est.	Parm.	Est.	Parm.	Est.	Parm.	Est.
α_1	-7.80	β_{18} ($CC_{6(4)}$)	-.41	β_{39} (JT_7)	-.15	β_{60} ($q82b_4$)	.53
α_2	-5.69	β_{19} ($CC_{7(4)}$)	0	β_{40} (JT_8)	-.28	β_{61} ($q82b_5$)	.14
α_3	-2.34	β_{20} ($CC_{1(5)}$)	-.11	β_{41} (JT_9)	-.62	β_{62} ($q82b_6$)	.23
α_4	-.045	β_{21} ($CC_{2(5)}$)	.092	β_{42} (JT_{10})	0	β_{63} ($q82b_7$)	0
β_1 (RC_1)	.11	β_{22} ($CC_{3(5)}$)	-1.63	β_{43} ($q79a_1$)	1.92	β_{64} ($q82d_1$)	1.27
β_2 (RC_2)	.59	β_{23} ($CC_{4(5)}$)	.11	β_{44} ($q79a_2$)	2.09	β_{65} ($q82d_2$)	.92
β_3 (RC_3)	1.23	β_{24} ($CC_{5(5)}$)	0	β_{45} ($q79a_3$)	1.43	β_{66} ($q82d_3$)	.67
β_4 (RC_4)	-.13	β_{25} ($CC_{1(6)}$)	-.13	β_{46} ($q79a_4$)	.84	β_{67} ($q82d_4$)	.77
β_5 (RC_5)	.59	β_{26} ($CC_{2(6)}$)	.34	β_{47} ($q79a_5$)	.58	β_{68} ($q82d_5$)	.44
β_6 (RC_6)	.40	β_{27} ($CC_{3(6)}$)	-.23	β_{48} ($q79a_6$)	.19	β_{69} ($q82d_6$)	.19
β_7 (RC_7)	0	β_{28} ($CC_{4(6)}$)	0	β_{49} ($q79a_7$)	0	β_{70} ($q82d_7$)	0
β_8 ($CC_{1(1)}$)	-.45	β_{29} ($CC_{1(7)}$)	.42	β_{50} ($q82a_1$)	2.68	β_{71} ($q82f_1$)	.85
β_9 ($CC_{2(1)}$)	-.60	β_{30} ($CC_{2(7)}$)	0	β_{51} ($q82a_2$)	.71	β_{72} ($q82f_2$)	1.69
β_{10} ($CC_{3(1)}$)	0	β_{31} (BC_1)	-.21	β_{52} ($q82a_3$)	1.05	β_{73} ($q82f_3$)	1.08
β_{11} ($CC_{1(2)}$)	0	β_{32} (BC_2)	0	β_{53} ($q82a_4$)	.89	β_{74} ($q82f_4$)	.69
β_{12} ($CC_{1(3)}$)	0	β_{33} (JT_1)	-.097	β_{54} ($q82a_5$)	.31	β_{75} ($q82f_5$)	.59
β_{13} ($CC_{1(4)}$)	.78	β_{34} (JT_2)	-.25	β_{55} ($q82a_6$)	.14	β_{76} ($q82f_6$)	.16
β_{14} ($CC_{2(4)}$)	-.53	β_{35} (JT_3)	-.20	β_{56} ($q82a_7$)	0	β_{77} ($q82f_7$)	0
β_{14} ($CC_{3(4)}$)	.83	β_{36} (JT_4)	-.48	β_{57} ($q82b_1$)	1.31		
β_{16} ($CC_{4(4)}$)	-.050	β_{37} (JT_5)	.11	β_{58} ($q82b_2$)	.81		
β_{17} ($CC_{5(4)}$)	.24	β_{38} (JT_6)	-.47	β_{59} ($q82b_3$)	.75		

Table 2. Maximum Likelihood Estimates of $\{\alpha_i\}_{i=1}^{4}$ and $\underset{\sim}{\beta}$

The section of the PROC LOGISTIC output entitled 'Type-3 Analysis of Effects' characterizes the statistical significance of the covariates through p-values obtained by referencing a Wald chi-square test statistic to a corresponding null chi-square distribution. Table 3 shows the chi-square tests and the corresponding p-values, and it is seen that all covariate groups are highly significant contributors in the model.

One way to assess model adequacy for multinomial logistic regression is to use the model to predict Y and then examine how well the predicted values match the true values of Y. Since the output of the model for each customer is an estimated probability distribution for Y, a natural predictor of Y is the mode of this distribution. We note that this predictor considers equal cost for all forms of prediction errors. More elaborate predictors could be derived by assuming a more complex cost model where, for example, the cost of predicting 5 when the actual value is 1 is higher than the cost of predicting 5 when the actual value is 4. Table 4, the so-called confusion matrix of the predictions, displays the cross classification of all 5056 customers based on their actual value of Y and the model-predicted value of Y.

Covariate Group	Degrees of Freedom	Wald Statistic	p-value
RC	6	41.2	< .0001
CC	16	40.9	< .01
BC	1	7.9	< .01
JT	9	43.7	< .0001
q79	6	84.8	< .0001
q82a	6	56.5	< .0001
q82b	6	34.4	< .0001
q82d	6	34.8	< .0001
q82f	6	39.9	< .0001

Table 3. Statistical Significance of Covariate Groups

Actual Y	Predicted Y					Total
	1	2	3	4	5	
1	3	2	7	4	4	20
2	3	8	48	22	7	88
3	2	3	342	486	133	966
4	0	0	126	1233	723	2082
5	0	0	39	705	1156	1900
Total	8	13	562	2450	2023	5056

Table 4. Confusion Matrix of Multinomial Logistic Regression Model

A perfect model would have a confusion matrix that is diagonal indicating the predicted value for each customer coincided identically with the true value. Consider the rows of Table 4 corresponding to Y=4 and Y=5. These two rows account for almost 80% of the customers in the sample. It can be seen that in both cases, the predicted value coincides with the actual value about 60% of the time. Neither of these two cases predicts Y=1 or Y=2, and only 4% of the time is Y=3 predicted. The mean values of the predicted Y when Y=4 and Y=5 are 4.28 and 4.59, respectively. The 7% positive bias for the case Y=4 is roughly offset by the 11.8% negative bias for the case Y=5.

Looking at the row of Table 4 corresponding to Y=3, we see that 86% of the time the predicted Y is within 1 of the actual Y. The mean value of the predicted Y is 3.77, indicating a 26% positive bias. Considering the rows corresponding to Y=1 and Y=2, where only about 2% of the customers reside, we see the model struggles to make accurate predictions, often over-estimating the actual value of Y. A hint as to the explanation for the noticeable over-estimation associated with the Y=1, Y=2 and Y=3 customers is revealed by examining their responses to the covariate questions. As just one example, the respective mean scores on question q79 ("Overall satisfaction with the service event") are 3.8, 4.1 and 5.2. It seems a relatively large number of customers that give a low response to Y are inclined to simultaneously give favorable responses to the covariate questions on the survey. Although this might be unexpected, it can possibly be explained by the fact that the covariate questions are relevant to the most recent service event whereas Y is based on a customer's cumulative experience.

Overall, Table 4 reflects significant lift afforded by the multinomial logistic regression model for predicting Y. For example, a model that utilized no covariate information would have a

confusion matrix whose rows were constant, summing to the row total. In sum, we feel the accuracy of the model is sufficient to learn something about what drives customers to give high responses to Y, though perhaps not sufficient to learn as much about what drives customers to give low responses to Y.

Figure 1 is a graphical display of the slopes for each of the customer satisfaction covariates. The larger the coefficient value, the more detrimental the response level is to NPS. The y-axis is therefore labeled as 'demerits.'

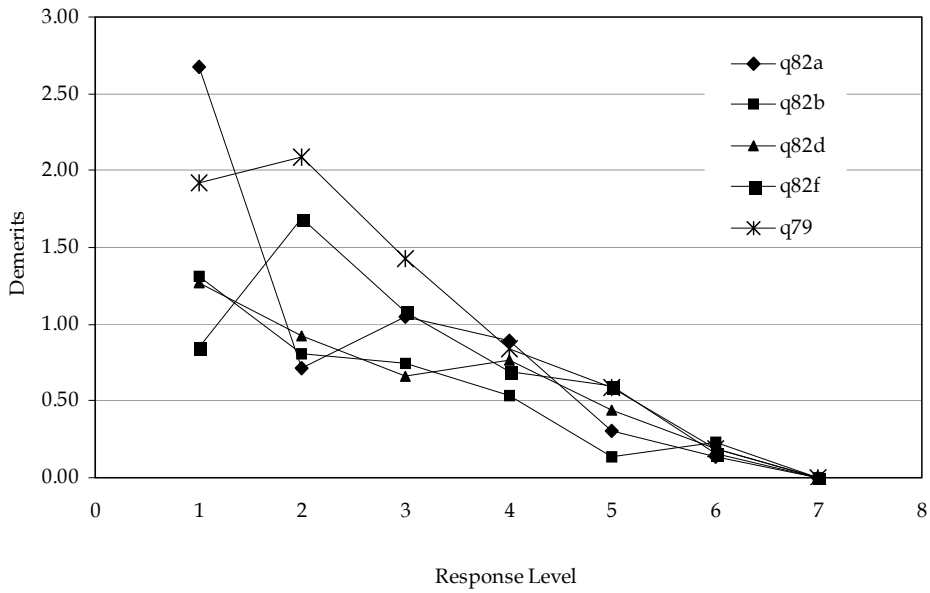


Fig. 1. MLEs of Slopes for 7-Point Likert Scale Customer Satisfaction Covariates

In view of the ordinal nature of the customer satisfaction covariates, the slopes, which represent the effect of the Likert scale levels, should decrease monotonically. That is, the penalty for a 'satisfied' covariate value should be less than or equal to that of a 'dissatisfied' covariate value. As such, it would be logical to have the estimated values of the slopes display the monotone decreasing trend as the response level of the covariates ascends. Figure 1 shows that the unconstrained MLEs for the slopes associated with the customer satisfaction covariates nearly satisfy the desired monotone property, but not exactly. The aberrations are due to data deficiencies or minor model inadequacies and can be resolved by using a constrained logistic regression model introduced in the next section.

3.3 Constrained logistic regression

Consider the situation where the *i*-th covariate is ordinal in nature, perhaps because it is measured on a *k*-point Likert scale. The CSH data is a good illustration of this situation, since all the customer satisfaction covariates are ordinal variables measured on 7-point Likert scale. Let the corresponding group of *k* slopes for this covariate be denoted by $\{\beta_{ij}\}_{j=1}^{k}$. In order to reflect the information that the covariates are ordered, it is quite natural

to impose the monotone constraint $\beta_{i1} \geq \beta_{i2} \cdots \geq \beta_{ik}$ onto the parameter space. Adding these constraints when finding the MLEs complicates the required maximization of the likelihoods in (1) and (3). In this section, however, we will show how this can be done using SAS with PROC NLP.

In order to simplify our use of PROC NLP, it is convenient to work with a full-rank parameterization of the logistic regression model. Because countries are nested within regions, a linear dependency exists between the dummy variables corresponding to regions and countries within regions. We can eliminate the linear dependency by removing region from the model and specifying country to be non-nested factor. The result of this model reparameterization is that instead of 6 degrees of freedom in the model for regions and 16 degrees of freedom for countries nested within regions, we equivalently have 22 degrees of freedom for countries. For the same purpose, we also redefine the dummy variable coding used for other categorical and ordinal covariates by using a full rank parameterization scheme. In particular, we use *k*-1 dummy variables (rather than *k*) to represent a *k*-level categorical or ordinal variable. With the full rank parameterization, the highest level of customer satisfaction has a slope parameter that is fixed to be 0. Lines 3-10 in the SAS code shown in Appendix B are used to set up the full rank parameterization of the logistic regression model.

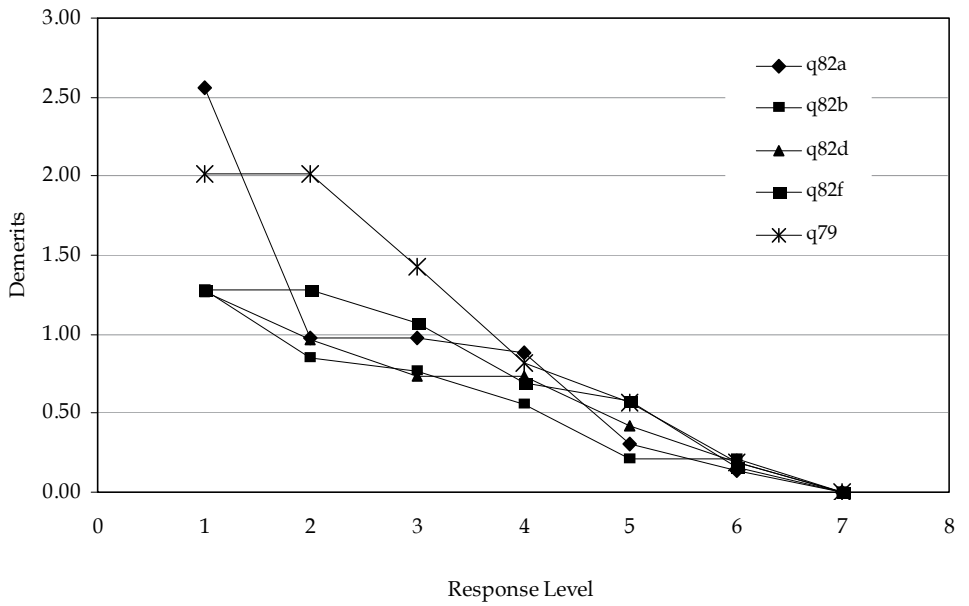


Fig. 2. Constrained MLEs of Slopes for 7-Point Likert Scale Customer Satisfaction Covariates

Beginning with line 12 in the SAS code, PROC NLP is used to derive the MLEs of the parameters under the constrained parameter space. The 'max' statement (line 13) indicates the objective function is the log-likelihood function of the model and that it is to be maximized. The maximization is carried out using a Newton-Raphson algorithm, and the 'parms' statement (line 14) specifies initial values for the intercept and slope parameters. The SAS variables bq*j*, ba*j*, bb*j*, bd*j* and bf*j* are used to symbolize the slope parameters corresponding to the *j*-th response level of the customer satisfaction covariates q79, q82a,

q82b, q82d and q82f. Similarly, bcc*j*, bbc*j*, and bj*j* are used to denote the slopes associated with different countries, business codes and job titles. The 'bounds' and 'lincon' statements (lines 15-21) jointly specify the monotone constraints associated with the intercept parameters and the slopes of the customer satisfaction covariates. Lines 22-29 define the log likelihood for each customer which, for the *i*-th customer, is given by

$$\text{loglik}_i(\underset{\sim}{\alpha}, \underset{\sim}{\beta}) = \sum_{j=1}^{5} Y_{ij} \log P(Y_i = j)$$

Covariate	MLE of Slope		Covariate	MLE of Slope	
	Unconstrained	Constrained		Unconstrained	Constrained
$q79a_1$	1.92	2.01	$q82d_1$	1.27	1.27
$q79a_2$	2.09	2.01	$q82d_2$	.92	.96
$q79a_3$	1.43	1.43	$q82d_3$	.67	.73
$q79a_4$	.84	.82	$q82d_4$	.77	.73
$q79a_5$	.58	.57	$q82d_5$	.44	.42
$q79a_6$	.19	.19	$q82d_6$	.19	.19
$q79a_7$	Structural 0	Implied 0	$q82d_7$	Structural 0	Implied 0
$q82a_1$	2.68	2.56	$q82f_1$	.85	1.28
$q82a_2$	.71	.98	$q82f_2$	1.69	1.28
$q82a_3$	1.05	.98	$q82f_3$	1.08	1.07
$q82a_4$	.89	.88	$q82f_4$	.69	.69
$q82a_5$	.31	.30	$q82f_5$	.59	.58
$q82a_6$	.14	.14	$q82f_6$	.16	.16
$q82a_7$	Structural 0	Implied 0	$q82f_7$	Structural 0	Implied 0
$q82b_1$	1.31	1.28			
$q82b_2$	.81	.85			
$q82b_3$	.75	.77			
$q82b_4$	.53	.56			
$q82b_5$	.14	.21			
$q82b_6$	.23	.21			
$q82b_7$	Structural 0	Implied 0			

Table 5. Unconstrained and Constrained Slope MLEs of Customer Satisfaction Covariates

Table 5 provides a side-by-side comparison of the constrained and unconstrained MLEs for the slopes of the customer satisfaction covariates, and Figure 2 is a plot that shows the monotone behavior of the constrained estimates. There is very little difference between the unconstrained and constrained MLEs for the demographic covariates. Recall that for the unconstrained MLEs, the zero for the slope of the last level of each covariate is a structural zero resulting from the non-full rank dummy variable coding used when fitting the model.

In the case of the constrained MLEs, the slopes of the last levels of the covariates are implied zeros resulting from the full-rank dummy variable coding used when fitting the model. Table 5 shows that incorporating the constraints do not lead to a substantial change in the estimated slopes. In an indirect way, this provides a sanity check of the proposed model. We will use the constrained estimates for the remainder of the case study.

3.4 Model utility

3.4.1 NPS for individual customers

The estimated coefficients of the model can be used to predict the distribution of Y for a customer with a given set of covariates as follows. Suppose a customer is from the second Country Code within the first Region Code, has the third Job Title and is associated with the second Business Code. These demographic covariates are coded as follows:

RC=(1,0,0,0,0,0,0)
CC=(0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
JT=(0,0,1,0,0,0,0,0,0,0)
BC=(0,1)

Suppose further that the customer gives sub-element scores for q79, q82a, q82b, q82d and q82f of 6, 6, 7, 5 and 6, respectively. These sub-element covariates are coded as follows:

q79=(0,0,0,0,0,1,0)
q82a=(0,0,0,0,0,1,0)
q82b=(0,0,0,0,0,0,1)
q82d=(0,0,0,0,1,0,0)
q82f=(0,0,0,0,0,1,0)

Combining these covariates into $\underset{\sim}{x}$, using the estimates of $\{\alpha_i\}_{i=1}^{4}$ and $\underset{\sim}{\beta}$ given in Table 2, and evaluating the equations in (2) gives the probability distribution for Y for this customer profile as:

$$Y \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ .00052 & .00377 & .10437 & .43860 & .45274 \end{pmatrix}$$

from which it follows that customers with this profile $\underset{\sim}{x}$ will have an (expected) NPS of 75.2%.

3.4.2 NPS for a population of customers

Consider now the sub-population of customers in the first Region Code. For this sub-population, the relative frequencies of the three Country Codes are 81.18%, 14.12% and 4.7%, respectively. The relative frequencies of the ten job titles are 11.76%, 17.65%, 12.94%, 18.82%, 27.06%, 3.53%, 0%, 4.71%, 2.35% and 1.18%. The relative frequencies of the two Business Codes are 82.35% and 17.65%, respectively. Thus, the demographic covariate vectors for this sub-population are:

$$\mathrm{RC} = (1,0,0,0,0,0,0)$$
$$\mathrm{CC} = (.8118,.1412,.0047,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$$
$$\mathrm{JT} = (.1176,.1765,.1294,.1882,.2706,.0353,0,.0471,.0235,.0118)$$
$$\mathrm{BC} = (.8235,.1765)\,.$$

Additionally, the distributions for the sub-element scores within the first Region Code define the sub-element covariate vectors as follows:

q79=(.0118,0,.0235,.0471,.3176,.2824,.3176)
q82a=(.0118,0,.0118,.0471,.2471,.4000,.2824)
q82b=(.0118,.0118,.0235,.1176,.2235,.3529,.2588)
q82d=(.0118,.0118,.0118,.0588,.2353,.3765,.2941)
q82f=(.0118,0,.0235,.0235,.1882,.4471,.3059)

Combining these covariates into $\underset{\sim}{x}$, using the estimates of $\{\alpha_i\}_{i=1}^{4}$ and $\underset{\sim}{\beta}$ given in Table 2, and evaluating the equations in (2) gives the probability distribution for Y for this population profile as:

$$Y \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ .00070 & .00503 & .13452 & .47771 & .38204 \end{pmatrix}$$

from which it follows that customers with this profile $\underset{\sim}{x}$ will have an (expected) NPS of 70.1%.

3.4.3 Targeted Pathways for Improving NPS

A purely empirical way to compute NPS is to use the observed distribution (based on all 5,056 survey responses) of Y for $\underset{\sim}{p}$ in the formula $\mathrm{NPS} = \sum_{i=1}^{5} w_i p_i$, and this yields 61.7%. Consider now filling out the covariate vector $\underset{\sim}{x}$ with the sample frequencies for the observed demographic covariates and with the observed sample distributions for the sub-element covariates. Using this $\underset{\sim}{x}$ with the model yields a predicted NPS of 65.7%. The close agreement between the data-based and model-based NPS scores is additional evidence that the model fits the data well, and it also instills confidence in using the model to explore "What If?" scenarios as outlined in Figure 3. Figure 3 defines sixteen "What If?" scenarios, labels them with brief descriptions, and then shows expected NPS score if the scenario is implemented. Table 6 contains a longer description of how each scenario was implemented. Each scenario can be evaluated on the basis of how much boost it gives to the expected NPS as well as the feasibility of establishing a company program that could make the hypothetical scenario real.

We illustrated potential pathways to improve the overall NPS score, but this can also be done with specific sub-populations in mind. For example, if the first region was under study, then one could simply adjust the demographic covariates as illustrated in section 3.4.2 before implementing scenarios adjustments.

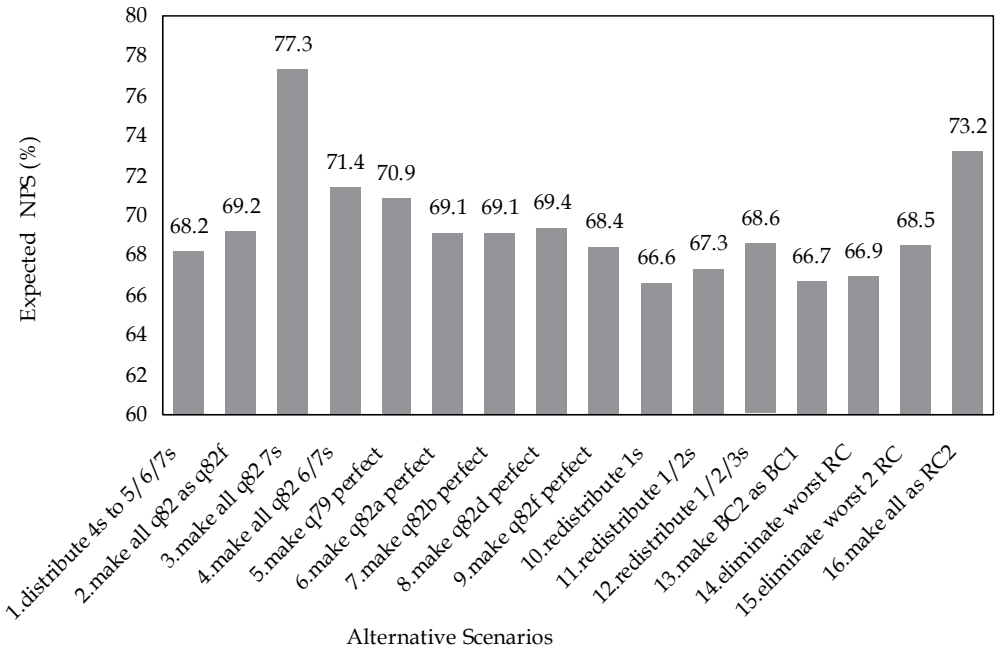


Fig. 3. Predicted NPS for Different Scenarios

Scenario	Brief Description
1	For each of q79, q82a, q82b, q82d and q82f, alter the distributions of their responses by reassigning the probability of a neutral response (4) equally to the probability of responses (5), (6) and (7)
2	Replace the response distribution for sub-elements q82a, q82b, and q82d with what was observed for q82f (which was the sub-element that had the most favorable response distribution)
3	Make the response distribution for each of q82a, q82b, q82d and q82f perfect by placing all the probability on response (7)
4	Improve the response distribution for each of q82a, q82b, q82d and q82f by placing all the probability equally on responses (6) and (7)
5	Improve the response distribution for q79 by placing all the probability on response (7)
6	Improve the response distribution for q82a by placing all the probability on response (7)
7	Improve the response distribution for q82b by placing all the probability on response (7)
8	Improve the response distribution for q82d by placing all the probability on response (7)
9	Improve the response distribution for q82f by placing all the probability on response (7)
10	Improve the response distribution for each of q79, q82a, q82b, q82d and q82f by distributing the probability of response (1) equally on responses (2)-(7)
11	Improve the response distribution for each of q79, q82a, q82b, q82d and q82f by distributing the sum of the probability of responses (1) and (2) equally on responses (3)-(7)
12	Improve the response distribution for each of q79, q82a, q82b, q82d and q82f by distributing the sum of the probability of responses (1), (2) and (3) equally on responses (4)-(7)
13	Simulate making Business Code 2 as good as Business Code 1 by setting BC=(1,0)
14	Improve the response distributions of q79, q82a, q82b, q82d, and q82f by replacing them by the average across the different Region Codes, excluding the worst Region Code
15	Improve the response distributions of q79, q82a, q82b, q82d, and q82f by replacing them by the average across the different Region Codes, excluding the two worst Region Codes
16	Improve the response distributions of q79, q82a, q82b, q82d, and q82f by replacing them all by the observed, respective, distributions for Region Code 2 (which was the region that had the most favorable response distribution)

Table 6. Implementation Detail for Each Scenario

4. Discussion

Alternative measures to NPS of customer advocacy include customer satisfaction (CSAT) and Customer Effort Score (CES) (Dixon et al., 2010). CES is measured on a 5-point scale and is intended to capture the effort required by a customer to resolve an issue through a contact-center or self-service channel. (Dixon et al., 2010) compared the predictive power of CSAT, NPS and CES on service customers' intention to do repeat business, increase their spending, and speak positively about the company. They concluded that CSAT was a relatively poor predictor, while CES was the strongest. NPS ranked in the middle.

The choice of which customer advocacy measure to use depends on many factors such as the type of company-to-customer relationship, the degree to which recommendations (for or against a company) influence a purchase decision, and whether the measures will be complemented by other customer feedback. To gain an in-depth understanding of customers' experiences and how to improve them may require multiple indicators. In the end, it is the action taken to drive improvements that customers value that is most critical.

Our case study validates the feasibility for using a multinomial logistic regression model as a means to identify key drivers of NPS, though it is clear that the same methodology could be employed with alternative measures of customer advocacy. Improvement teams at CSH have used this model to prioritize projects relative to their expected impacts on NPS. A novel aspect of our model development was the implementation of monotone constraints on the slope parameters of the ordinal covariates. Our illustrative SAS code showing how to impose the constraints on the maximum likelihood estimates should be of significant help to practitioners interested in doing the same thing.

5. Appendix A

```
data indata;
infile 'C:\CarestreamHealth\indata.txt';
input RC CC BC JT Y q79 q82a q82b q82d q82f;
run;

proc logistic data=indata;
class RC CC BC JT
      q79 q82a q82b q82d q82f/param=glm;
model Y = RC CC(RC) BC JT
      q70 q79 q82a q82b q82d q82f;
run;
```

6. Appendix B

```
data indata;
set indata;
array cc{23} cc1-cc23; do i=1 to 23; if CC=i then cc{i}=1; else cc{i}=0;end;
if BC=1 then bc1=1;else bc1=0;
array jt{9} jt1-jt9; do i=1 to 9; if JT=i then jt{i}=1; else jt{i}=0;end;
array q{6} q1-q6; do i=1 to 6; if q79=i then q{i}=1; else q{i}=0;end;
array a{6} a1-a6;  do i=1 to 6; if q82a=i then a{i}=1; else a{i}=0;end;
```

```
array b{6} b1-b6; do i=1 to 6; if q82b=i then b{i}=1; else b{i}=0;end;
array d{6} d1-d6; do i=1 to 6; if q82d=i then d{i}=1; else d{i}=0;end;
array f{6}  f1-f6;  do i=1 to 6; if q82f=i then f{i}=1; else f{i}=0;end;
run;

proc nlp data=indata;
max loglik;
parms alp1=-7, alp2=-5, alp3=-2, alp4=-1, bcc1-bcc10=0, bcc12-bcc23=0, bbc1=0, bj1-bj9=0,  bq1-bq6=0, ba1-ba6=0, bb1-bb6=0,bd1-bd6=0, bf1-bf6=0;
bounds 0 <= bq6,0 <= ba6,0 <= bb6,0 <= bd6,0 <= bf6;
lincon 0<=alp4-alp3, 0<=alp3-alp2, 0<=alp2-alp1;
lincon 0<= bq5-bq6, 0<= bq4-bq5, 0<= bq3-bq4, 0<= bq2-bq3, 0<= bq1-bq2;
lincon 0<= ba5-ba6, 0<= ba4-ba5, 0<= ba3-ba4, 0<= ba2-ba3, 0<= ba1-ba2;
lincon 0<= bb5-bb6, 0<= bb4-bb5, 0<= bb3-bb4, 0<= bb2-bb3, 0<= bb1-bb2;
lincon 0<= bd5-bd6, 0<= bd4-bd5, 0<= bd3-bd4, 0<= bd2-bd3, 0<= bd1-bd2;
lincon 0<= bf5-bf6, 0<= bf4-bf5, 0<= bf3-bf4, 0<= bf2-bf3, 0<= bf1-bf2;
tp=cc1*bcc1+cc2*bcc2+cc3*bcc3+cc4*bcc4+cc5*bcc5+cc6*bcc6+cc7*bcc7+cc8*bcc8+cc9*bcc9+cc10*bcc10+cc12*bcc12+cc13*bcc13+cc14*bcc14+cc15*bcc15+cc16*bcc16+cc17*bcc17+cc18*bcc18+cc19*bcc19+cc20*bcc20+cc21*bcc21+cc22*bcc22+cc23*bcc23+bc1*bbc1+jt1*bj1+jt2*bj2+jt3*bj3+jt4*bj4+jt5*bj5+jt6*bj6+jt7*bj7+jt8*bj8+jt9*bj9+q1*bq1+q2*bq2+q3*bq3+q4*bq4+q5*bq5+q6*bq6+a1*ba1+a2*ba2+a3*ba3+a4*ba4+a5*ba5+a6*ba6+b1*bb1+b2*bb2+b3*bb3+b4*bb4+b5*bb5+b6*bb6+d1*bd1+d2*bd2+d3*bd3+d4*bd4+d5*bd5+d6*bd6+f1*bf1++f2*bf2+f3*bf3+f4*bf4+f5*bf5+f6*bf6;
pi1=exp(alp1+tp)/(1+exp(alp1+tp));pi2=exp(alp2+tp)/(1+exp(alp2+tp));
pi3=exp(alp3+tp)/(1+exp(alp3+tp));pi4=exp(alp4+tp)/(1+exp(alp4+tp));
if Y=1 then loglik=log(pi1);
if Y=2 then loglik=log(pi2-pi1);
if Y=3 then loglik=log(pi3-pi2);
if Y=4 then loglik=log(pi4-pi3);
if Y=5 then loglik=log(1-pi4);
run;
```

7. References

Allen, D. R. & Rao, T. R. N. (2000). *Analysis of Customer Satisfaction Data*, ASQ Quality Press, 978-0873894531, Milwaukee.

Dixon, M., Freeman, K. and Toman, N. (2010). *Stop Trying to Delight Your Customers*, *Harvard Business Review*, July-August, http://hbr.org/magazine.

Hayes, B. E. (2008). *Measuring Customer Satisfaction and Loyalty, 3rd edition*, ASQ Quality Press, 978-0873897433, Milwaukee.

Keiningham, T.; Cooil, B.; Andreassen, T. & Aksoy, L. (2007). A Longitudinal Examination of Net Promoter and Firm Revenue Growth, *Journal of Marketing,* Vol.71, No.3, (July 2007) 39-51, 0022-2429.

Reichheld, F. (2003). The One Number You Need to Grow, *Harvard Business Review*, Vol. 81, No.12, (December 2003) 46-54.

Vavra, T. G. (1997), *Improving Your Measurement of Customer Satisfaction: Your Guide to Creating, Conducting, Analyzing and Reporting Customer Satisfaction Measurement Programs*, ASQ Quality Press, 978-0873894050, Milwaukee.

9

The Declarative Framework Approach to Decision Support for Constrained Search Problems

Paweł Sitek and Jarosław Wikarek
Technical University of Kielce
Poland

1. Introduction

Decision making supported by task-oriented software tools plays a pivotal role in modern enterprises; the commercially available ERP (Enterprise Resource Planning) systems are not able to respond in an interactive on-line/real-time mode.
A new generation of DSS (Decision Support System) that enable a fast prototyping of constrained search problems (e.g. layout planning, production routing, batch-sizing, scheduling, transportations, distributions, flow in supply chain,) is needed. In that context for constrained search problems, the CLP (Constraint Logic Programming) (Apt et al, 2007) techniques allowing declarative representation of a decision making problem provide quite an attractive alternative. Constrained search problems are usually characterized by many types of constraints, which make them unstructured and difficult to solve (NP-complete). Traditional mathematical programming approaches are deficient because their representation of constraints is artificial (using 0-1 variables).
This paper discusses declarative framework for decision support in outbound logistic (Fig.1) of SCM (Supply Chain Management) (Simchi-Levi et al., 2003) (Douglas et al., 2000) as an example of use. It will focus on potential areas and processes where decision support or optimization is reasonable, and it will also introduce the concept and an outline of decision support system structures for the SCM in the form of an additional layer of information. The solutions are developed not to substitute, but to be integrated with SCM and with the broader sense Enterprise Resource Planning (ERP) solutions.

2. Supply chain management

Increasing competition in today's global market and the heightened expectations of customers have forced enterprises to consider their supply chains more carefully. A supply chain (SC) can be considered as a network of stages that represent functionalities (including suppliers, manufacturers, distributors, and retailers) that must be provided to convert raw materials into the specified end-products and deliver these end-products to retailers or customers (Simchi-Levi et al., 2003). A supply chain system (SCS) is usually composed of a series of organizations and/or independent companies. A supply chain system is the set of approaches utilized to efficiently integrate suppliers, manufacturers, warehouses,

distribution centers, retailers and ultimately the customers so that merchandise is produced and distributed at the right quantities to the right locations and at right times, in order to minimize the total system cost while satisfying the service level requirements. The main operational activities of the supply chain system include:

1. Raw materials: sales forecasting, inventory planning, and purchasing, transportation between suppliers and manufacturers – Suppliers.
2. Work-in-process: processing and managing efficiently the inventories of partially completed products inside the manufacturing plants – Manufacturers.
3. Finished goods: warehousing of finished goods inventory, servicing the customers, and transporting materials among the wholesalers, retailers, and customers. These parts reflect the basic movement of goods, services, and information through an integrated system-Distributors, Retailers, Customers.

A supply chain is characterized by a forward flow of goods and a backward flow of information (Fig.1). Commonly, a supply chain is compromised of two basic business processes:

- material management
- physical distribution

Increasingly, the management of multiple relationships across the supply chain is being referred to as supply chain management (SCM) (Douglas et al., 2000).

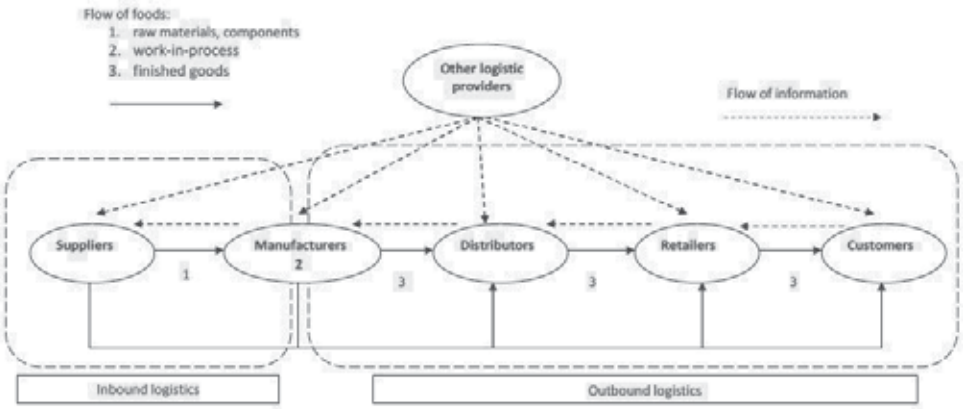


Fig. 1. The supply chain process.

The field of supply chain management uses an increasing set of tools and techniques to effectively coordinate the decision-making process.

2.1 Integration in SCM

Supply-chain management is one of the leading business process re-engineering, cost-saving and revenue enhancement strategies in use today. Effective integration of a supply chain can save money, simultaneously improving customer service and reducing inventories etc.. Supply chains have been more or less integrated to some extent – albeit on a low scale – as a whole, or in parts. Integration, if done at all, has been mostly done in patches throughout the supply chain. In many cases, this has been driven more by the need to survive and improvise, than by the willingness to improve and advance further. Integration can be

carried out for different aspects of a company's supply chain. According to (Shapiro, 2001), there are four dimensions to integration.

- **Functional integration**: It relates to the different functions performed by the organization, for example, purchasing, location, manufacturing, warehousing, and transportation.
- **Spatial integration**: This type of integration is done over a target group of supply chain entities – vendors, manufacturing facilities, warehouses, and markets.
- **Hierarchical integration**: It refers to integration of the overlapping decisions in the strategic, tactical, and operational planning horizons. For example, it is important to strategically locate the manufacturing facilities with respect to the vendors to help in optimizing the supply policies.
- **Enterprise integration**: It emphasizes the importance of strategic and tactical planning decisions like, integration of supply chain management with demand management and financial management to maximize the revenues and also to increase long-term returns on investments.

2.2 Decision support in SCM

Several issues need to be considered while building a decision support system (DSS) for managing the supply chain. A manufacturer will find offers to sell the materials they need as well as offers to purchase goods they produce. All this information will supplement the data already in the firm's ERP (Enterprise Resource Planning) system, which includes inventory, customer orders, and the state of the manufacturing schedule, distribution etc.. Taken together, these data can enable much better decisions as to what products to make, what raw materials to purchase, what orders to accept, how to distribute etc.. This paper will focus on logistic where decision support or optimization is very reasonable. A logistic network should maximize profits and provide a least total cost system, while also achieving desired customer service level. The models of outbound logistics are often mixed-integer programming (MIP) formulation that seek to achieve the following:

- Inventory optimization and reduction at the warehouse (the factory warehouse) or/and distribution center;
- Load optimization for transportation from the factory warehouse to the delivery locations (i.e., DCs, bins, plants, and direct-delivery (DD) customers, Distribution Centers(DC)),while allowing direct (plant-to-store) and interplant shipments;
- Flow optimization throughout the entire outbound supply chain.

A number of quantitative models use mixed-integer programming (MIP) to solve the supply chain optimization problems. One of the first attempts was done by Geoffrion and Graves [4], where a MIP model (Frühwirth & Slim, 2003) (Vanderbei, 2008) (Schrijver, 1998) (Cornelis et al., 2006) was formulated for the multicommodity location problem. This seminal research involved the determination of distribution center (DC) locations, their capacities, customer zones and transportation flow patterns for all commodities. A solution to the location portion of the problem was presented, based on Bender's Decomposition (BD). The transportation portion of the problem is decoupled into a separate classical transportation problem for each commodity. Their approach shows a high degree of effectiveness and advantage of using BD over branch-and-bound. The technique has been applied on a real problem to test its performance. However, the computational requirements and technical resources required for its implementation make it a difficult choice in classical MIP tools.

3. The concept of decision support system structures – declarative approach

While imperative models specify exactly how things have to be done, declarative approaches only focus on the logic that governs the interplay of actions in the process by describing (1) the activities that can be performed, as well as (2) constraints prohibiting undesired behavior. Imperative models take an 'inside-to-outside' approach by requiring all execution alternatives to be explicitly specified in the model. Declarative models, in turn, take an 'outside-to-inside' approach: constraints implicitly specify execution alternatives as all alternatives have to satisfy the constraints (Pesic, 2008).
The concept of decision support system for managing supply chain can be regarded as an additional information layer, which contains both elements of the structure and functional information.

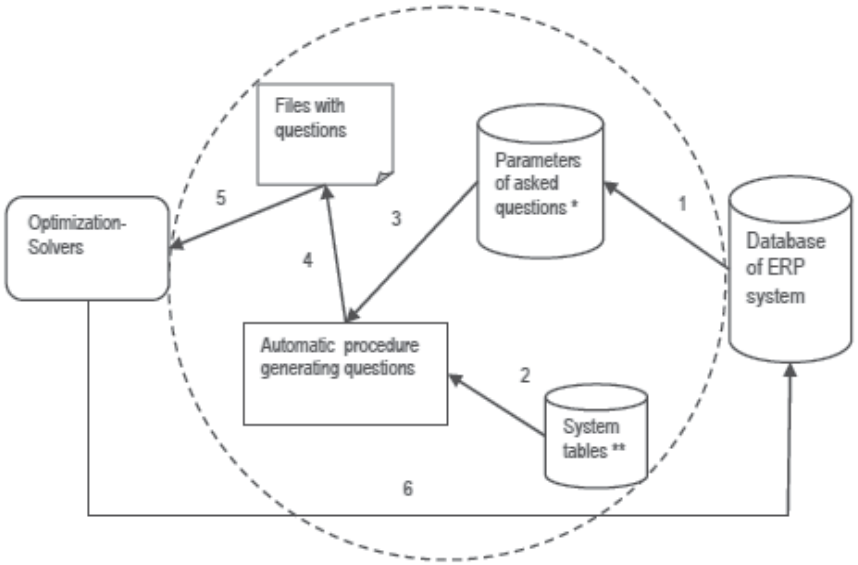


Fig. 2. The simplified schema of the basic structures and functionality for decision support system as an additional information layer

Decisions in the presented system are the answers to specific questions. Questions are implemented as predicates in CLP. The contribution of this concept is largely automated implementation of questions as predicates CLP by appropriate entries in the database.
The parts of the proposed database structure for the decision support system are presented in Table 1, Table 2 and ERD (Entity Relationship Diagram)-Fig.3, Fig.4, Fig.5. The basic functional structure, components of the system and method of operation are shown in Fig.1. Ask a question launches a sequence of events. The most important of them subject to the chronology and numbers (Fig.2) have been presented below.

1. Mapping-completion of the data structures of questions based on the ERP database (Table 2).
2. Loading information on the structure and functions of questions from the system tables (Table 1)(Fig.3).

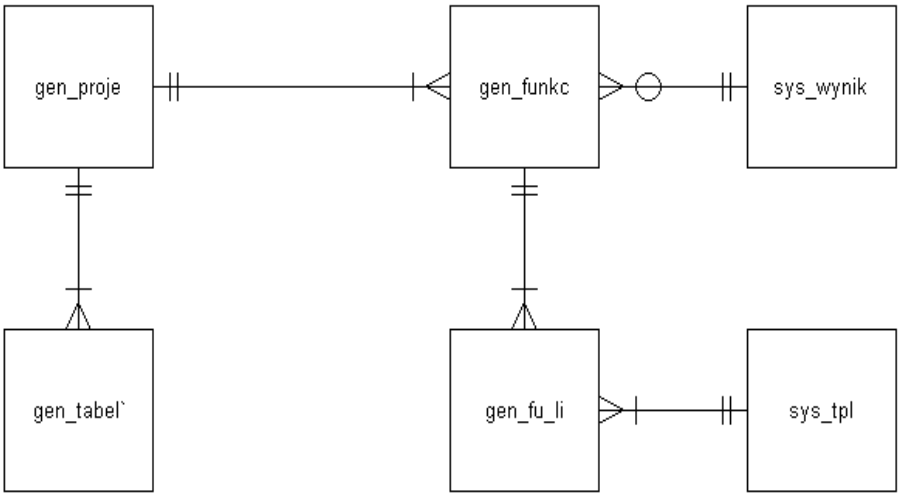


Fig. 3. The part of ERD (simplified Martin's notation) diagram for the system tables.

Table name *description*	**Column name**	**Key**	**Column description**
sys_typl -*Types of line*	kd_idlin	pk	ID line type
	id_nazwa		The name of the line type
sys_wynik -*Results*	kd_wynik	pk	ID result
	id_opis_		Description of the type of result
gen_proje *Projects*	kd_idpro	pk	ID project
	id_nazwa		The name of the project
	id_skrot		Short name
	id_opis_		Description of the project
gen_funkc- *Functions during the handling of the project*	kd_idpro	pk(fk)	ID project
	kd_idfun	pk	ID function
	fd_wynik	fk	Id result
	id_opis_		Description of the function
gen_fu_li - *The function codes during the operation of the project*	kd_idpro	pk(fk)	ID project
	kd_idfun	pk(fk)	ID function
	kd_id_ko	pk	The line number
	fd_idlin	fk	ID line type
	id_dana_		The line code
gen_tabel -*Tables of the project*	kd_idpro	pk(fk)	ID project
	kd_id_ta	pk	The name of the table in the project.
	id_id_al		Table name
	id_opis_		Table description

Table 1. System tables - the part of the additional information layer.

3. Loading data for questions from questions parameters tables.
4. Generating questions in the form of text files in the correct format/meta-language/ of the optimization program.
5. Sending files with questions to the optimization program.
6. Recording the results obtained/decisions/in the database.

(question=decision making model)
The functionality of the layer mapped by the dotted line in Fig.2.
Decision can be interpreted as a response to the properly formulated question that meets the existing constraints.
As an illustrative example of using DSS (the questions in the form of the CLP predicates), the optimization in distribution centers has been presented. Philosophy and structure of the proposed decision support system may have a place for all areas of SCM, ERP (Moon, 2007) etc.

4. Optimization in distribution centers

Centralization of distribution reduces the number of transactions compared to the number of transactions without a central distribution. The supplier does not need to send parcel to multiple recipients, but sends one to a dedicated distribution center. Similarly, the recipient does not need to take many items from many suppliers, but receives the bulk of the load from the distribution center.
Reducing the number of transactions between suppliers and customers helps reduce the average time of delivery and the costs of these transactions. The centralized distribution in the form of distribution centers results in an incomparably greater opportunity to optimize resources, processes, costs, etc.
Potential areas of decision support and optimization for the exemplified distribution center depends on the decision level:
Strategic level – (number and location of the distribution center depots, the choice of product groups, the territorial constraints of the area serviced by the distribution center, transportation problems etc.);
Tactical level - (structure and size of fleet vehicles, periodic change of route plans, etc.);
Operational level - (completion of contracts , optimization of loading, dynamic route planning, inventory management, high storage, allocation of pallets etc.).
It should be noted that these effects are achieved at the cost of what is happening inside the distribution center. It is there where the received goods must be unloaded, reassembled and uploaded for transport to customers. Therefore the issue of cost optimization is not obvious for every lot size or for any batch of goods.

4.1 Problem of allocation of pallets

Problems of optimization and decision support in a distribution center play a key role at the operational level, where decisions must be taken daily or even several times a day.
The method and quality of such decisions have the greatest impact on the operation of the distribution center and as a result on its effectiveness.
It is at the operational level where rapid and effective assistance should be provided to operators who manage the warehouse activities, including palletizing and forwarding. One of several problems that must be solved every day in distribution centers (e.g. wholesale food, liquor, etc.) is the problem of allocation of pallets.

For example, at the distribution center of FMCG (Fast Moving Consumer Goods), pallets distribution process for the end customer begins with the storage document created by the Sales Department on the basis of a client's order.

The storage document contains general information about the contract, including the division into pallets. Based on this information picking and packing is done. The arrangement of individual goods in the storage document warehouse is optimal against the AVG route. Then the pallets are stored in the forwarding area of the warehouse. The allocation of pallets to each route and then the way of loading the pallets on trucks are decided by dispatchers on the basis of their experience. The method of loading (loading order) pallets on the truck is the deterministic process that depends on the route. But the allocation of pallets to the route and the appropriate type of truck is not the deterministic process.

In the problem of allocation of pallets, the main asked question is: What is the minimum cost of supply pallets? Answer to this question in this case the solution of mathematical model allows for answers to other questions. These are important questions from a practical point of view. The most important ones are: Which route will be operated? What is the number of courses assigned to each route? What type of trucks will be used? How much will the supplies for each point?

Symbol	Description
	Indices
j	delivery point $j \in J=\{1,..,N\}$
i	route $i \in I=\{1,..,M\}$
l	truck type $l \in L=\{1,..,O\}$
	Input Parameters
Z_j	demand for pallets at delivery point j (j=1..N).
D_{ji}	$\begin{cases} 1 & \text{if delivery point } j \text{ belongs to route } i \\ 0 & \text{otherwise} \end{cases}$ for (j=1..N), (i=1..M)
K_l	capacity of truck l (pallets) (l=1..O)
U_l	feasible number of routes for truck type l (l=1..O).
W_l	feasible length of route for truck type l (l=1..O).
C_{li}	the cost of route i (i=1..M) by the truck type l (l=1..O)
E_{li}	the time course of truck type l (l = 1..O) on the route i (i=1..M)
$B_{j,i,l}$	the cost of delivery to delivery point j (j =1..N) on the route i (i = 1..M) by truck type l (l =1..O). (Variable depending on the number of pallets).
	Decision Variables (Integer)
$X_{j,i,l}$	part of the demand for pallets for the delivery point j (j=1..N), implemented on the route i (i=1..M) supported by the truck type l (l=1..O)
$Y_{i,l}$	number of courses truck type l (l =1..O) on the route i (i =1..M)

Table 2. The indices, input parameters and the decision variables for optimization model.

A mathematical model for the main asked question has been formulated in the form of Mixed Integer Programming (MIP) problem (Schrijver, 1998). The cost of transport, which as a result of optimization is minimized has been adopted as an objective function. Decision variables of the model (X_{jil}) are the variables determining the number of pallets to be delivered to the delivery point of a route chosen by the truck.
Constraints of the mathematical model (1) to (6) can be interpreted as follows. Constraint (1) ensures that every requirement is met, that is, that each delivery point receives as many pallets of goods as they ordered. Constraint (2) ensures the fulfillment of the transfer for it restricts quantitatively the loading on the truck to its potential capacity and the number of courses. A further constraint (3) ensures that the number of courses for the given type of truck is not exceeded. Constraint (4) enforces the allocation of vehicles to these routes where cargo has already been allocated in the form of a certain number of pallets, i.e. the value of decision variable X_{jil} is different from zero. Constraint (6) refers to a timely delivery. Constraint (5) ensures that all the decision variables are restricted to be integers. The presented model also takes into account the dynamic optimization of routes. To this end a delivery points that belong to each route has been introduced, and the decision variable determining the volume of demand X_{jil} depends on specific points of supply.
The indices, input parameters and the decision variables for this model have been shown in Table 2.
Objective Function – minimize transportation costs:

$$\sum_{i=1}^{M}\sum_{l=1}^{O} Y_{i,l} * C_{i,l} + \sum_{j=1}^{N}\sum_{i=1}^{M}\sum_{l=1}^{O} X_{j,i,l} * B_{j,i,l}$$

Subject to:

$$\sum_{i=1}^{M}\sum_{l=1}^{O} X_{j,i,l} * D_{j,i} = Z_j \quad \text{for j} = 1..\text{N} \tag{1}$$

$$\sum_{j=1}^{N}\sum_{i=1}^{M} X_{j,i,l} \le U_l * K_l \quad \text{for l} = 1..\text{O} \tag{2}$$

$$\sum_{i=1}^{M} Y_{i,l} \le K_l * Y_{i,l} \quad \text{for } l = 1..\text{O}, \tag{3}$$

$$\sum_{j=1}^{N} X_{i,j,l} \le K_l * Y_{i,l} \quad \text{for i} = 1..\text{M}, l = 1..\text{O}, \tag{4}$$

$$X_{i,j,l} \in C \quad \text{for j} = 1..\text{N}, \text{i} = 1..\text{M}, l = 1..\text{O},$$

$$Y_{i,l} \in C \text{ for i} = 1..\text{M}, l = 1..\text{O}, \tag{5}$$

$$\sum_{i=1}^{M} E_{i,l} Y_{l,l} \le W_l \text{ for } l = 1..\text{O} \tag{6}$$

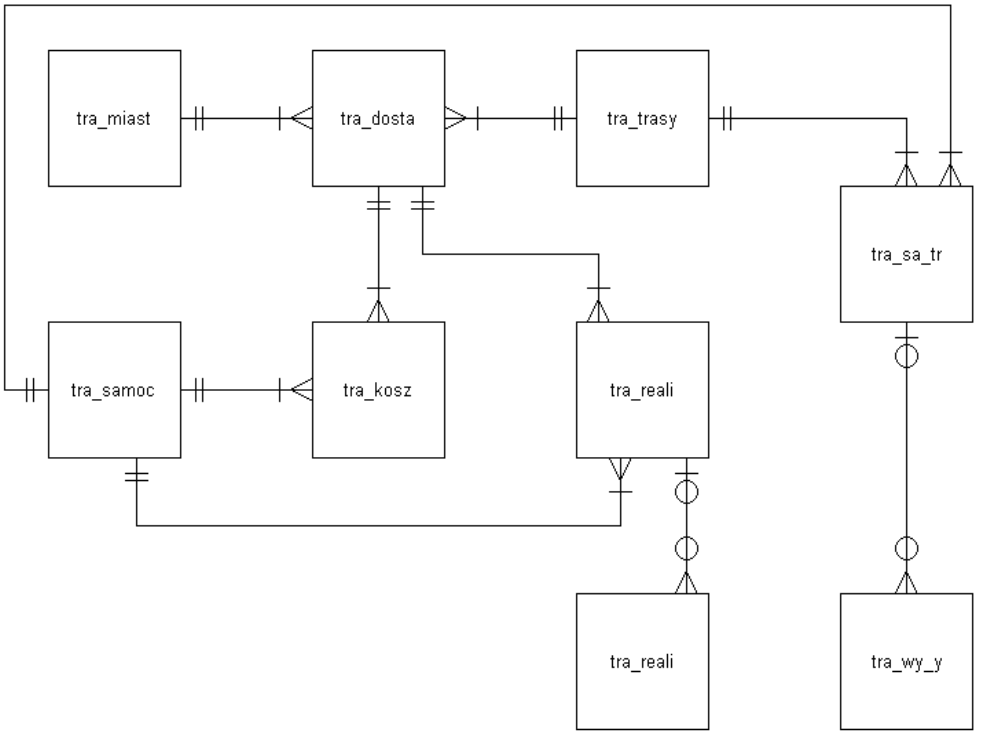


Fig. 4. The part of ERD (simplified Martin's notation) diagram for questions parameters tables.

Table name *description*	**Symbol**	**Column name**	**Key**	**Column description**
tra_miast *delivery points*	j	kd_idmia	pk	ID delivery point
	naz	id_nazwa		The name of the point
	z	id_zapot		The demand for pallets
tra_trasy *Routes*	i	kd_idtra	pk	ID route
	naz	id_nazwa		The name of the route
tra_dosta *delivery points along the route*	j	kd_idmia	pk(fk)	ID delivery point
	i	kd_idtra	pk(fk)	ID route
tra_samoch *truck types*	l	kd_idsam	pk	ID type of truck
	k	id_poj_k		The capacity of truck type (in pallets)
	u	id_kur_u		Feasible number of curses
	w	id_cza_w		Feasible duration of the courses
tra_sa_tr *allocation of trucks to the routes*	i	kd_idtra	pk(fk)	ID route
	l	kd_idsam	pk(fk)	ID type of truck
	c	id_kos_c		The cost of route by the truck type

	e	id_cza_e		The time course of truck type on the route
	y	iw_kursy		The number of curses of truck type (solution)
tra_reali *number of pallets*	j	kd_idmia	pk (fk)	ID delivery point
	i	kd_idtra	pk (fk)	ID route
	l	kd_idsam	pk (fk)	ID type of truck
	X	iw_dosts		The number of pallets (solution)
tra_koszt *the cost of delivery*	j	kd_idmia	pk(fk)	ID delivery point
	i	kd_idtra	pk(fk)	ID route
	l	kd_idsam	pk(fk)	ID type of truck
	b	id_koszt		the cost of delivery

Table 3. Questions parameters tables (problem of allocation of pallets)-the part of the additional information layer

Practical application of decision support system is shown for the problem of allocation of pallets in regional FMCG distribution center. In this example, asked question (minimize transportation costs) has been implemented as a model of the MIP. The Question parameters tables for the above problem and the relationships between the tables have been presented in Table 3 and Fig 4.

The full MIP model as a CLP predicate, generated based on the discussed information layer in the form of a text file accepted by the ECLIPSE -CLP (Apt et al, 2007) package is shown in Fig.5.

```
:- module(dane).
:- lib(fd).
:- lib(fd_global).
:- lib(edge_finder3).
:- lib(branch_and_bound).
:- local struct(zadanie(start,czas,wymaga,ludzie)).
:- export
dane/12
.
dane(Z, D, K, U, W, C, E, X, Y,Xopr,Yopr,Razem):-
  Z = [25,10,15,10,20,30,25,15],
  D = [[1,1,1,0,0],[1,1,0,1,0],[1,1,0,1,0],[0,0,1,1,1],[0,0,1,1,1],
       [1,0,1,0,1],[0,1,1,1,0],[1,1,1,0,0]],
  K = [10,6],
  U = [12,12],
  W = [58,58],
  C = [10,8,12,9,16,13,12,9,7,5],
  E = [ [ 5,6,8,6,6 ], [ 5,6,8,6,6 ]],
  B = [8,6,8,6,8,6,0,0,0,0,6,4,6,4,0,0,6,4,0,0,6,5,6,5,0,0,6,5,0,0,0,0,0,0,
       5,5,6,6,5,5,0,0,0,0,6,4,6,4,6,4,6,5,0,0,6,5,0,0,6,5,0,0,8,6,8,6,8,6,
       0,0,8,6,8,7,8,7,0,0,0,0],
X=[[[Xp111,Xp112],[Xp121,Xp122],[Xp131,Xp132],[Xp141,Xp142],[Xp151,Xp152]],
   [[Xp211,Xp212],[Xp221,Xp222],[Xp231,Xp232],[Xp241,Xp242],[Xp251,Xp252]],
   [[Xp311,Xp312],[Xp321,Xp322],[Xp331,Xp332],[Xp341,Xp342],[Xp351,Xp352]],
   [[Xp411,Xp412],[Xp421,Xp422],[Xp431,Xp432],[Xp441,Xp442],[Xp451,Xp452]],
   [[Xp511,Xp512],[Xp521,Xp522],[Xp531,Xp532],[Xp541,Xp542],[Xp551,Xp552]],
   [[Xp611,Xp612],[Xp621,Xp622],[Xp631,Xp632],[Xp641,Xp642],[Xp651,Xp652]],
```

```
   [[Xp711,Xp712],[Xp721,Xp722],[Xp731,Xp732],[Xp741,Xp742],[Xp751,Xp752]],
   [[Xp811,Xp812],[Xp821,Xp822],[Xp831,Xp832],[Xp841,Xp842],[Xp851,Xp852]]
  ],
Xopr=[[[Xp111,Xp211,Xp311,Xp411,Xp511,Xp611,Xp711,Xp811],
       [Xp121,Xp221,Xp321,Xp421,Xp521,Xp621,Xp721,Xp821],
       [Xp131,Xp231,Xp331,Xp431,Xp531,Xp631,Xp731,Xp831],
       [Xp141,Xp241,Xp341,Xp441,Xp541,Xp641,Xp741,Xp841],
       [Xp151,Xp251,Xp351,Xp451,Xp551,Xp651,Xp751,Xp851]],
      [[Xp112,Xp212,Xp312,Xp412,Xp512,Xp612,Xp712,Xp812],
       [Xp122,Xp222,Xp322,Xp422,Xp522,Xp622,Xp722,Xp822],
       [Xp132,Xp232,Xp332,Xp432,Xp532,Xp632,Xp732,Xp832],
       [Xp142,Xp242,Xp342,Xp442,Xp542,Xp642,Xp742,Xp842],
       [Xp152,Xp252,Xp352,Xp452,Xp552,Xp652,Xp752,Xp852]]],
(foreach(X1,X), foreach(Z1, Z) do (foreach(X11,X1), param(Z1) do
   (foreach(X111,X11), param(Z1) do X111 #<= Z1, X111 #>= 0 ) ) ),
Y=[[Yp11,Yp12],[Yp21,Yp22],[Yp31,Yp32],[Yp41,Yp42],[Yp51,Yp52]],
Yopr=[[Yp11,Yp21,Yp31,Yp41,Yp51],[Yp12,Yp22,Yp32,Yp42,Yp52]],
(foreach(U1,U), foreach(Y1,Yopr) do (foreach(Y2,Y1), param(U1) do
   Y2 #>= 0,  Y2 #<= U1 ) ),
flatten(Y,Ypo), flatten(C,Cpo),
(foreach(C1,Cpo),foreach(Y1,Ypo),foreach(Fy1,Fun_y) do  Fy1 #=C1 * Y1 ),
flatten(X,Xpo),
(foreach(B1,B),foreach(X1,Xpo),foreach(Fx1,Fun_x) do  Fx1 #=B1 * X1),
sumlist(Fun_y,Razem1), sumlist(Fun_x,Razem2), Razem #= Razem1+Razem2
```

Fig. 5.a. The MIP model (input data) for the problem of allocation of pallets- automatically generated (Eclipse-CLP).

```
:- module(funkcje).
:- use_module(dane).
:- lib(fd).
:- lib(fd_global).
:- lib(fd_search).
:- lib(edge_finder3).
:- lib(branch_and_bound).
:- lib(probing_for_scheduling).
:- local struct(zadanie(start,czas,wymaga,ludzie)).
:- local struct(nazwa(napis)).
:- export
opc/0
.
og_d1(X,D,Z):-
 (foreach(X1,X), foreach(D1,D), foreach(Z1,Z) do
   (foreach(X2,X1), foreach(D2,D1), param(Z1) do
     (foreach(X3,X2), param(D2), param(Z1) do  X3 #<= D2*Z1 ) ) ).
og_1(X,Z):-
 (foreach(X1,X), foreach(Z1,Z) do flatten(X1,X2), sumlist(X2,Z1) ).
og_2(K,U,Xopr):-
 (foreach(K1,K), foreach(U1,U), foreach(X1,Xopr) do
    flatten(X1,X2), sumlist(X2,X3), X3 #<= U1*K1 ).
og_3(U,Yopr):-
 (foreach(U1,U),foreach(Y1,Yopr) do sumlist(Y1,Y2), Y2 #<=U1 ).
ogr_4(Xopr,K,Yopr):-
 (foreach(K1,K), foreach(Y1,Yopr), foreach(X1,Xopr) do
  (foreach(Y2,Y1), param(K1), foreach(X2,X1) do
```

```
   sumlist(X2,X3), X3 #<= K1 * Y2 ) ).
ogr_6(E,Yopr,W):-
(foreach(W1,W), foreach(Y1,Yopr), foreach(E1,E) do
 (foreach(E2,E1),foreach(Y2,Y1), foreach(C1,C) do
  C1 #= E2 * Y2 ), sumlist(C,C3), C3 #<= W1 ).

s1(L_L1,L_L2):-
  flatten(L_L1,L1),  flatten(L_L2,L2),
  append(L2,L1,L3),
  search(L2,0, most_constrained,indomain_split,complete,[]),
  search(L1,0, most_constrained,indomain_split,complete,[]).
zapis(X,Y):-

 open("w_1.txt",write,Sp_1), flatten(X,X1),
 (foreach(X2,X1), param(Sp_1) do write(Sp_1,X2), writeln(Sp_1,'  ') ),
  writeln(Sp_1,' '), close(Sp_1),
 open("w_2.txt",write,Sp_2), flatten(Y,Y1),
 (foreach(Y2,Y1), param(Sp_2) do write(Sp_2,Y2), writeln(Sp_2,'  ') ),
 writeln(Sp_2,' '),
 close(Sp_2).
opc:-

dane(Z, D, K, U, W, C, E, X, Y,Xopr,Yopr,Razem),
og_d1(X,D,Z),
og_1(X,Z),
og_2(K,U,Xopr),
og_3(U,Yopr),
ogr_4(Xopr,K,Yopr),
bb_min(s1(Yopr,Xopr),Razem, bb_options with [strategy: step]),
zapis(X,Y).
```

Fig.5.b. The MIP model (constraints) for the problem of allocation of pallets- automatically generated (Eclipse-CLP).

Computational experiments were performed using the data from Table 4a. The calculations are made for three different sets of parameters U, W (examples 1, 2, 3). Results were as follows: Example1 Cost = 1054, Example2 Cost = 1028, Example3 Cost = 1050. The differences are caused by changes in the parameters U and W. The values of decision variables (X, Y) corresponding to the optimal solutions are presented in Table 4b. There is a solution of Example2 from the practice in this table. This is a feasible solution. What is the most interesting when comparing the two solutions of Example2, that address the optimal supply of pallets is possible using a smaller number of courses. This information is important because of the costs and organization of work in other areas such as on the daily demand for trucks, drivers, etc.

tra_miast	
j	Z
1	25
2	10
3	15
4	10
5	20

tra_dosta							
j	i	j	i	j	i	j	i
1	1	3	1	5	3	7	2
1	2	3	2	5	4	7	3
1	3	3	4	5	5	7	4
2	1	4	3	6	1	8	1
2	2	4	4	6	3	8	2

tra_sa_tr			
i	l	C	E
1	1	10	5
1	2	8	5
2	1	12	6
2	2	9	6
3	1	16	8

tra_samoc			
l	K	U	W
Example1			
1	10	10	80
2	6	10	80
Examplet2			
1	10	12	80

6	30
7	25
8	15

2	4	4	5	6	5	8	3

3	2	13	8
4	1	12	6
4	2	9	6
5	1	7	6
5	2	5	6

2	6	12	80
Example3			
1	10	12	58
2	6	12	58

tra_koszt															
j	**i**	**l**	**B**	**j**	**i**	**l**	**B**	**j**	**i**	**l**	**B**	**j**	**i**	**l**	**B**
1	1	1	8	3	1	1	6	5	3	1	6	7	2	1	8
1	1	2	6	3	1	2	5	5	3	2	4	7	2	2	6
1	2	1	8	3	2	1	6	5	4	1	6	7	3	1	8
1	2	2	6	3	2	2	5	5	4	2	4	7	3	2	6
1	3	1	8	3	4	1	6	5	5	1	6	7	4	1	8
1	3	2	6	3	4	2	5	5	5	2	4	7	4	2	6
2	1	1	6	4	3	1	5	6	1	1	6	8	1	1	8
2	1	2	4	4	3	2	5	6	1	2	5	8	1	2	6
2	2	1	6	4	4	1	6	6	3	1	6	8	2	1	8
2	2	2	4	4	4	2	6	6	3	2	5	8	2	2	7
2	4	1	6	4	5	1	5	6	5	1	6	8	3	1	8
2	4	2	5	4	5	2	5	6	5	2	5	8	3	2	7

Table 4.a. Data for computational examples

Example1 fc =1054-optimal			
j	**i**	**l**	**X**
1	1	1	11
1	1	2	14
2	1	2	10
3	1	1	14
3	4	1	1
4	5	1	10
5	4	2	2
5	5	2	18
6	5	1	30
7	4	1	9
7	4	2	16
8	1	1	15
	i	**l**	**Y**
	1	1	4
	1	2	4
	4	1	1
	4	2	3
	5	1	4
	5	2	3

Example2 fc=1026-optimal			
j	**i**	**l**	**X**
1	1	2	25
2	1	2	10
3	1	1	14
3	4	1	1
4	5	1	10
5	5	1	2
5	5	2	18
6	1	1	2
6	5	1	28
7	2	2	6
7	4	1	19
8	1	1	2
8	1	2	13
	i	**l**	**Y**
	1	1	2
	1	2	8
	2	2	1
	4	1	2
	5	1	4
	5	2	3

Example2 fc=1119-feasible			
j	**i**	**l**	**X**
1	1	2	25
2	1	2	5
2	4	2	5
3	1	1	15
4	3	1	10
5	3	1	2
5	5	2	18
6	1	1	23
6	3	1	7
7	3	1	6
7	4	2	19
8	3	1	15
	i	**l**	**Y**
	1	1	4
	1	2	5
	2	1	4
	3	1	4
	4	2	4
	5	2	3

Example3 fc=1050-optimal			
j	**I**	**l**	**X**
1	1	2	25
2	1	1	2
2	1	2	8
3	1	1	15
4	5	1	10
5	5	1	8
5	5	2	12
6	5	1	30
7	4	1	19
7	4	2	6
8	1	2	15
	i	**l**	**Y**
	1	1	2
	1	2	8
	4	1	2
	4	2	1
	5	1	5
	5	2	2

Table 4.b. Solutions for computational examples

4.2 Multi-stage transportation problem

The standard is determining the delivery of homogenous materials (commodity) from *n* suppliers to *m* purchasers. Multi-stage transportation problem, which in addition to suppliers and customers, there are still intermediate points such as distribution centers (Fig.2). Where the set W = {W_1, W_2, .. W_n} specifies the suppliers/manufacturers/, the set O = {O_1, O_2, .. O_m} purchasers (delivery points) and set P = {P_1, .. P_k} intermediaries such as warehouses, distribution centers, logistic warehouses.

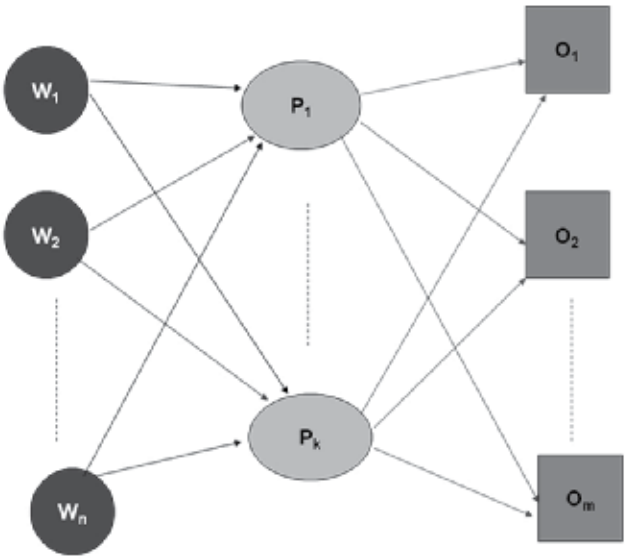


Fig. 6. The network for multi-stage transportation problem

Symbol	Description
Indices	
j	delivery point (j=1..M)
i	facility/factory (i=1..N)
s	distribution center (s=1..E)
N	the number of factories
M	the number of delivery points
E	the number of distribution centers
Input Parameters	
C_j	production cost for the factory i (i=1..N).
W_i	production capacity for factory i (i=1..N).
Z_j	demand for delivery point j (j=1..M)
A_{is}	the cost of delivery from factory i to distribution center s (i=1..N) (s=1..E)
G_{sj}	the cost of delivery from distribution center s to delivery point j (s=1..E) (j=1..M)
Decision Variables	
X_{is}	the part of the delivery from factory i to distribution center s
Y_{sj}	the part of delivery from distribution center s to delivery point j

Table 5. The indices, input parameters and the decision variables for optimization model.

A mathematical model of optimization for multi-stage transportation problem has been formulated as a linear programming problem of minimizing the objective function which represents the total cost of transport and production under constraints (1) to (5). Constraint (1) ensures that every requirement is met, that is, that each distribution center receives delivery from the factory due to the production capacity. The fulfillment of the delivery for delivery point enforces constraint (2). Balancing the distribution centers, i.e. the supply and shipping ensures constraint (3). Constraints (4), (5) are the standard constraints for the problem of integer linear programming.

Objective Function – minimize transportation and production costs:

$$\sum_{i=1}^{N}\sum_{s=1}^{E} A_{i,s} * X_{i,s} + \sum_{s=1}^{E}\sum_{j=1}^{M} G_{s,j} * Y_{s,j} + \sum_{i=1}^{N}(C_i * \sum_{s=1}^{E} X_{i,s})$$

Subject to:

$$\sum_{s=1}^{E} X_{i,s} \le W_i \text{ for i = 1..N} \tag{1}$$

$$\sum_{s=1}^{E} Y_{s,j} \ge Z_j \text{ for j = 1..M} \tag{2}$$

$$\sum_{i=1}^{N} X_{i,s} = \sum_{j=1}^{M} Y_{s,j} \text{ for s = 1..E} \tag{3}$$

$$X_{i,s} \ge 0 \text{ for i = 1..N ,s = 1..E, integer} \tag{4}$$

$$Y_{s,j} \ge 0 \text{ for s = 1..E, j = 1..M, integer} \tag{5}$$

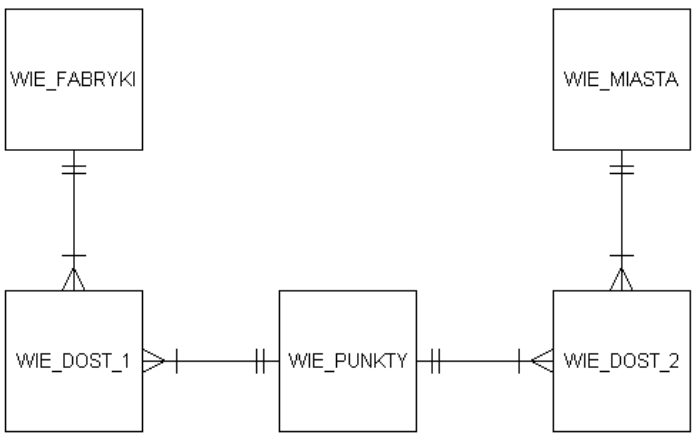


Fig. 7. The part of ERD (simplified Martin's notation) diagram for questions parameters tables

In this example, asked question (minimize transportation and production costs) has been implemented also as a model of the MIP. The question parameters tables for the above problem and the relationships between these tables have been presented in Table 6a, Table 6b and Fig. 7.
The full MIP model as a CLP predicate, generated based on the discussed information layer in the form of a text file accepted by the ECLIPSE – CLP package is shown in Fig.8a and Fig.8b.

```
:- module(dane_l_p).
:- lib(fd).
:- lib(fd_global).
:- lib(edge_finder3).
:- lib(branch_and_bound).
:- export
dane/8
.
dane(Z,C,W,X,Xt,Yt,Y,Razem):-
  Z = [4, 3, 2, 2, 3, 4],
  C = [180, 200, 190 ],
  W = [12, 10, 10],
  A = [30, 20, 90, 40, 80, 60],
  G = [30, 20, 40, 70, 50, 60, 90, 40, 50, 30, 40, 40],
  X = [ [X11, X12], [X21, X22], [X31, X32] ],
  Xt =[ [X11, X21, X31], [X12, X22, X32] ],
  Y = [ [Y11, Y12, Y13, Y14, Y15, Y16], [Y21, Y22, Y23, Y24, Y25, Y26] ],
  Yt =[[Y11,Y21],[Y12,Y22],[Y13,Y23],[Y14,Y24],[Y15, Y25],[Y16,Y26] ],
flatten(Y,Ypo),
flatten(X,Xpo),
(foreach(X1,X),
  foreach(W1,W) do
  (foreach(X11,X1), param(W1) do
    X11 #>= 0,
    X11 #<= W1 ) ),
(foreach(Y1,Yt),
  foreach(Z1,Z) do
  (foreach(Y11,Y1), param(Z1) do
    Y11 #>= 0,
    Y11 #<= Z1 ) ),
(foreach(X1,Xpo),
  foreach(A1,A),
   foreach(Sx1,Sum_x) do
   Sx1 #=X1 * A1 ),
sumlist(Sum_x,Razem1),
(foreach(Y1,Ypo),
   foreach(G1,G),
   foreach(Sy1,Sum_y) do
   Sy1 #=Y1 * G1 ),
sumlist(Sum_y,Razem2),
(foreach(Xk1,X),
  foreach(C1,C),
   foreach(Sk1,Sum_k) do
    sumlist(Xk1,Raz_s),
    Sk1 #=Raz_s * C1 ),
sumlist(Sum_k,Razem3),
Razem #= Razem1+Razem2+Razem3
.
```

Fig. 8.a. The MIP model (input data) for the multi-stage transportation problem-automatically generated (Eclipse-CLP).

```
:- module(funkcje_l_p).
:- use_module(dane_l_p).
:- lib(fd).
:- lib(fd_global).
:- lib(fd_search).
:- lib(edge_finder3).
:- lib(branch_and_bound).
:- lib(probing_for_scheduling).
:- local struct(zadanie(start,czas,wymaga,ludzie)).
:- local struct(nazwa(napis)).
:- export
opc/0
.
og_1(X,W):-
 (foreach(X1,X),
   foreach(W1,W) do
    sumlist(X1,X_sum), X_sum #<= W1 ).
og_2(Yt,Z):-
 (foreach(Y1,Yt),
   foreach(Z1,Z) do
    sumlist(Y1,Z1) ).
og_3(Y,Xt):-
 (foreach(Y1,Y),
   foreach(X1,Xt) do
    sumlist(Y1,Y_sum),
    sumlist(X1,X_sum), Y_sum #= X_sum ).
s1(L_L1,L_L2):-
  flatten(L_L1,L1),
  flatten(L_L2,L2),
  append(L2,L1,L3),
  search(L3,0, most_constrained,indomain_min,complete,[]).
zapis(X,Y):-
  open("w_1.txt",write,Sp_1),
  flatten(X,X1),
  (foreach(X2,X1), param(Sp_1) do
   write(Sp_1,X2),
   writeln(Sp_1,'  ') ),
  writeln(Sp_1,' '),
  close(Sp_1),
  open("w_2.txt",write,Sp_2),
  flatten(Y,Y1),
  (foreach(Y2,Y1), param(Sp_2) do
   write(Sp_2,Y2),
writeln(Sp_2,'  ') ),
  writeln(Sp_2,' '),
  close(Sp_2).
opc:-
 dane(Z,C,W,X,Xt,Yt,Y,Razem),
 og_1(X,W),
 og_2(Yt,Z),
 og_3(Y,Xt),
 bb_min(s1(X,Yt),Razem, bb_options with [strategy: step]),
 zapis(X,Yt).
```

Fig. 8.b. The MIP model (constraints) for the multi-stage transportation problem-automatically generated (Eclipse-CLP)

Example P1-parameter A

(i) \ (s)	1	2	C_i	W_i
1	30	20	180	600
2	90	40	200	500
3	80	60	190	500

Example P1-parameter G

(s) \ (j)	1	2	3	4	5	6
1	30	20	40	70	50	60
2	90	40	50	30	40	40
Zj	200	150	100	100	150	200

Example P2- parametr A

(i) \ (s)	1	2	3	4	5	C_i	W_i
1	30	40	30	40	30	180	600
2	90	40	40	50	40	200	500
3	80	60	70	70	50	190	500
4	30	20	30	40	30	150	300

Example P2-parametr G

(s) \ (j)	1	2	3	4	5	6
1	30	20	40	70	50	60
2	90	40	50	30	40	40
3	20	50	60	60	90	60
4	30	30	50	70	60	60
5	90	40	50	30	40	40
Zj	300	250	300	220	150	200

Table 6.a. Data for computational examples

Example 1 fc = 26500						Example2 fc = 345400					
i	s	X	s	j	Y	i	s	X	s	j	Y
1	1	350	1	1	200	1	1	550	1	2	250
1	2	250	1	2	150	1	3	50	1	3	300
2	2	300	2	3	100	2	3	250	2	4	100
			2	4	100	2	5	250	2	6	200
			2	5	150	3	5	20	3	1	300
			2	6	200	4	2	300	5	4	120
									5	5	150

Table 6.b. Solutions for computational examples

Computational experiments were performed using the data from Table 6a. The calculations are made for two examples (different number of distribution centers, factories). Results were as follows: Example1 Cost = 26500, Example2 Cost = 345400. The

results obtained make it possible to find answers to additional questions: Which distribution centers are the most used? Which plants are the largest suppliers? Are there unnecessary distribution centers?

5. Conclusions

The paper presents a general concept for the decision support system for the outbound logistic in supply chain. The unique contribution of the presented concept in the form of an additional layer of information is flexibility. It can be used to implement any questions as developed constraint logic programming. It proposes a very flexible and versatile means of automatic generation of decision-making models based on the contents of the database.

It should be noted that the quality of the proposed decision-making models, which correspond to the respective entries to the system tables depends on the quality of generated decision. It is worth noting that the entries in the system tables for the type of decision-making model are executed only once. Ready-made questions (decision making models) are longer automatically generated even if you change the data in the database ERP system.

6. References

Apt K.R, Wallace M.G. (2007). *Constraint Logic Programming Using ECL^{i}PS^e*,ISBN-13978-0-521-88628-6, Cambridge.

Chapters in a print book

Cornelis, Roos. Terlaky Tamás. Vial Jean-Philippe. (2006). *Interior Point Methods for Linear Optimization.* Second Edition, ISBN 0387263780, Springer-Verlag.

Douglas, M. Lambert, Martha. Cooper C. (2000) *Issues in Supply Chain Management* in *Industrial Marketing Management*, ISSN 0019-8501, *Volume 29, Issue 1*, pp 65-83.

Frühwirth, Thom., Slim, Abdennadher., (2003). *Essentials of constraint programming*, ISBN 978-3-540-67623-2 ,Springer.

Geoffrion, A.M. Graves G.W. (1974). *Multicommodity distribution system design by Bender's decomposition* in *Management Science 20 (5*), ISSN 0025-1909, pp. 822-844.

Moon Y.B. (2007). *Enterprise Resource Planning (ERP): a review of the literature*, in *J.Management and Enterprise Development*, ISSN 1468-4330,Vol. 4, No. 3, pp. 235–264.

Pesic, M. (2008) *Constraint-Based Workflow Management Systems: Shifting Control to Users*. PhD thesis, ISBN 978-90-386-1319-2, Eindhoven University of Technology.

Schrijver, A. (1998). *Theory of Linear and Integer Programming.* ISBN 0-471-98232-6, John Wiley & sons.

Shapiro, J.F. (2001). *Modeling the Supply Chain*, ISBN 978-0-534-37741 Duxbury Press.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi E. (2003). *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies.* McGraw-Hill, ISBN 978-0-07-119896-7, New York.

Vanderbei, R. J. (2008). *Linear Programming: Foundations and Extensions.* 3rd ed., International Series in Operations Research & Management Science, Vol. 114, ISBN 978-0-387-74387-5, Springer Verlag.

10

Generic Model Base Design for Decision Support Systems in Revenue Management: Applications to Hotel and Health Care Sectors

Miguel Gutiérrez and Alfonso Durán
Área de Ingeniería de Organización, Universidad Carlos III de Madrid
Spain

1. Introduction

The sustainable development of an increasingly service-based economy requires procedures for the efficient allocation, to the various existing user classes, of non-storable service infrastructures with essentially fixed costs and whose value potential is dilapidated if not utilized. That decision is often taken in phases; for instance, assigning a hospital wing to a specific use (e.g. surgery or radiology) is a long-term decision, given the high refurbishing costs. Similarly, strategic decisions have to be made regarding the distribution of hotel rooms in single, double, suites... In a second decision stage, operational decisions will be taken, such as the individual patients or customers to whom those infrastructures will be assigned.

The aforementioned problem of infrastructure allocation is frequently addressed by Revenue Management (RM) techniques—also known as Yield Management—. RM techniques were initially developed to deal with strategies regarding the offering and allocation of flight seats. According to Zaki (2000) "the main objective of RM is to sell the right seat to the right customer at the right time for the right price to maximize the profit." In a more general approach, RM is the process of reacting or anticipating to the consumer behaviour, so that the revenue is optimized. It implies the use of dynamic forecasting models, the allocation of perishable resources/services to the diverse price categories and channels, as well as the price to charge for every category-channel (Talluri & Van Ryzin, 2004).

To accomplish these tasks, RM uses optimization heuristics, whose relative effectiveness depends strongly on the characteristics of the demand, such as predictability, differences in sensitivity to the price between the different segments of clients and temporary pattern of evolution of the relative weight of each segment of clients when the date of execution approaches.

On the other hand, RM algorithms reflect (and therefore are contingent upon) the design of the business process whose optimization they must support (in this case, infrastructure assignment) in three related ways:

- The specific traits of the business process, such as whether overbooking is allowed or not and whether two different prices can be offered simultaneously through two different channels or not. It would also encompass such aspects as alternative

distribution channels through which rooms can be reserved, quota allocation and reallocation to each channel, cancellation procedures and penalizations, pricing process (prices that can only increase as the execution/ arrival date approaches vs. prices that can go either up or down).

- The decision variable(s) to be optimized. In a hotel reservation process, the decision to be taken might be when (how long before the actual stay date) to switch from a lower price to a higher price, or alternatively consider this anticipation as a fixed parameter and try to optimize which price to switch to.
- The business objective(s) to achieve. In the hotel sector it will typically be profit; given its essentially fixed costs, this translates into managing/ maximizing revenue, hence the "Revenue Management" term. In other service sectors the optimization objective is not so clear, as it is the case of the Health Care sector. Nordgren (2009) discusses the complexity of the multi-faceted concept of value in this sector and thus the number of objectives that should be simultaneously taken into account when trying to find a match between the service providers and the potential users.

Although potential benefits to be gained from the use of advanced RM techniques to guide the allocation of infrastructures are substantial, in many cases merely applying the existing algorithms adds limited value. In its thorough literature review on RM research, Chiang *et al*. (2007) highlight the resulting complexity and difficulty of developing and implementing RM projects, to the point of suggesting that these projects are viewed as strategic, as opposed to tactical activities. There is a barrier to the adoption of RM algorithms, derived from its dependency on the concrete design of the business process. Each service organization uses a different business process design for infrastructure allocations (e.g., the pricing and reservation process in a hotel). These organizations do not want to restrict themselves to predetermined business process designs just to use a specific algorithm. Since the appropriate algorithm is contingent on that design, organizations must tailor their algorithms or redevelop them from scratch. Besides, given the currently prevailing non-systematic approach to algorithm development, this adaptation requirement, both initially and whenever the business process is redesigned, imposes a stiff hindrance, particularly to the SME, and also limits its adaptability to changing market conditions.

The work presented here intends to overcome that barrier, by taking advantage of the flexibility offered by a generic modeling approach to design the model base component of a Revenue Management-based Decision Support System (DSS) aiming at the efficient allocation of the abovementioned non-storable service infrastructures.

The Model Base can be identified as the key component in RM DSS according to the definition provided by Sprague. In his classical DSS framework, Sprague distinguishes three main components in a DSS (Sprague, 1980): database, model base and software system. The software system in turn is composed of three components: the DBMS (Database Management Software), which in conjunction with the database constitutes the data subsystem; the MBMS (Model Base Management Software), which in conjunction with the model base constitutes the model subsystem; and the dialog generation and management software which forms the system user interface. The model subsystem is within the scope of a DSS what is frequently referred to as a Model Management System (MMS), which according to Muhanna & Pick (1994) can be defined with wide agreement as "a software system that facilitates the development, storage, manipulation, control, and effective utilization of models in an organization."

We can thus characterize DSS in RM problems as *Model-Driven DSS* following the framework of Power (2002). This author presents different DSS classifications and proposes a framework of DSS that includes five basic types of DSS characterized by their dominant component, user groups, purpose and enabling technology: Data-Driven, Model-Driven, Knowledge-Driven, Document-Driven and Communications-Driven. With respect to Model-Driven DSS, Power (2002) outlines that "Model-Driven DSS emphasize access to and manipulation of a model." (Power 2002)

In the next section we present the proposed generic modeling approach consisting in a hierarchy of levels of abstraction, which provides the conceptual key to design a flexible model base component of a RM DSS. The remaining sections of the chapter first describe the model associated with each of the levels of the abovementioned hierarchy and show the potential of this approach with applications to specific cases taken from the quite different in nature Hotel and Health Care sectors. One of the reasons for including the Health Sector in this analysis is to show the applicability of the approach in intrinsically multiobjective/ multicriteria settings (Gupta & Denton 2008; Nordgren 2009).

2. Hierarchical modeling approach

As stated earlier, the key question to address in order to surmount the barrier to the successful utilization of DSS in RM problems is the intrinsic linkage between the algorithms and the specific design of the business process. Our approach to tackle this barrier stems from a hierarchical generic modeling approach of the business processes. A *business process model* defines the way a particular infrastructure element is assigned to one of its potential uses. In the example of a hotel, that would be the design of the room reservation process, considering aspects like alternative distribution channels through which rooms can be reserved, pricing process (prices that can only increase as the execution/arrival date approaches vs. prices that can go either up or down), etc.

In order to explain the proposed modeling hierarchy, we establish a parallelism with a simple example of a mathematical model as can be seen in Fig.1. There are three hierarchical modeling levels:

- The **business process metamodel** level corresponds to the highest abstraction layer. The generic model of the business processes related to infrastructure assignment issues is defined, providing the basic elements and their relationships, so that, through instantiation, the set of possible models is derived. In the example of the mathematical model, in this level we would find a metamodel describing the building blocks of a mathematical equation, such as variables and operators, as well as the definition of the possible ways to combine these elements to constitute the two members of a mathematical equation.
- The **business process model** level encompasses the direct instances of the metamodel defined in the former level. In the mathematical example, in this level we would find generic equations such as that represented in the Figure: $a - b = c^2$, a, b, and c being natural numbers.
- The **business process instance** level emerges as the instantiation of the former level, in the sense that each process instance will originate from a process model by means of assigning specific values to a subset of its generic elements. The instances will become feasibility or optimization problems, with a defined objective in the latter case. The

subset of valued-assigned elements will constitute the subset of parameters of the model, whereas the non-assigned elements will form the subset of variables of the thus defined feasibility/optimization problem. In the mathematical example, we will obtain different equations with the assignment of values to a subset of the model elements (a, b, c).

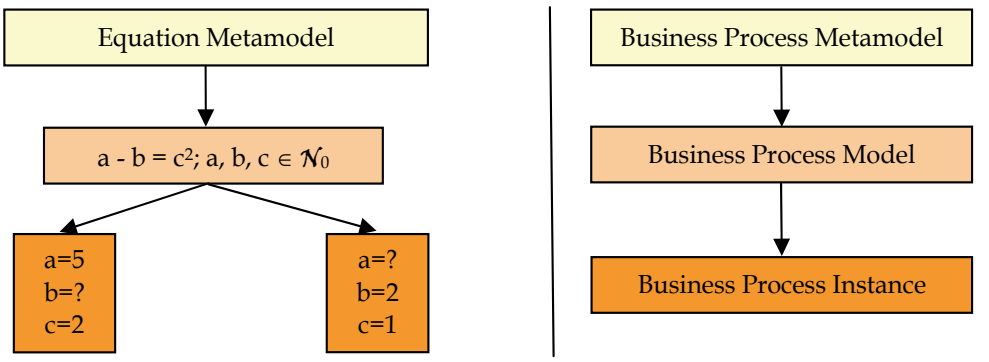


Fig. 1. Generic modeling hierarchy

Throughout the modeling process of the proposed hierarchy we make use of the standard UML, Unified Modeling Language (OMG, 2010). UML is an ongoing project developed by the OMG, Object Management Group, with the initial objective of standardizing and unifying the proliferation of object oriented modeling languages developed in the late 80s and early 90s (Fowler, 2004). According to OMG's description, UML supports the specification, visualization and documentation of software system models, including their structure and design, even though UML can also be used for business process modeling and modeling of other non-software systems too (OMG, 2011). The choice of UML derives from the confluence of these two traits: its applicability to business process modeling together with its software systems orientation (Eriksson & Penker, 2000) and the unifying standard nature with which the language was born, that has facilitated its growing usage and acceptance.

In the following sections we present the UML class diagrams that describe each of the levels of the business process hierarchy. Class Diagrams evolved from the Entity Relationship Diagrams (ERD). They describe the types of elements found in a system (entities) and the structural relationships among them. Entities are represented as square boxes and the relationships as lines that join two participating entities. There are two basic types of relationships: Association, in which participating entities are at the same level, and Generalization, graphically depicted as a line terminated in a triangle, in which one entity is a subtype or specialization of the other ("is-a-kind-of" relationship). Relationships may include a cardinality, or degree of multiplicity with which an entity participates in a relationship; it is denoted by the corresponding number, or by an asterisk to indicate any multiplicity. The default value for cardinality is one. Additionally, the role that an entity plays in a relationship might be annotated next to the entity.

3. Business process metamodel

The UML model that corresponds to the first abstraction layer is shown in Fig. 2. A *business process model* defines how a particular type of infrastructure is assigned to one of its potential

uses. The proposed metamodel, as the model of the process models, aims to support the whole spectrum of problem definitions intended to be addressed.

As shown in the class diagram, Infrastructure *Access Types* are defined first, as follows: each *Customer Segment Type* accesses through a *Channel Type* a certain *Infrastructure Type*. In the most general case, all *Customer Segment Types* would be able to access all *Infrastructure Types* through all *Channel Types*; the existence of restrictions in the access to some *Customer Segment Types*, or access to certain infrastructures through some *Channel Types*, will lead to a whole set of problems to be addressed.

This problem base is reflected in the instances of the entity *Infrastructure Access Type*. In turn, each of these instances might be assigned different *Value Types*. In commercial infrastructure allocation problems, these *Value Types* will be the tariff types, e.g. reduced tariff, normal tariff, premium tariff; offer price, list price; high season, low season... The assignment of a *Value Type* to an *Infrastructure Access Type* defines a generic *Allocation Type*. Each *Allocation Type* will apply during a time interval, which will be defined by a *Generic Start Time* and a *Generic End Time*. These times are handled in the generic model layer in abstract form, without taking a specific value or time. Even if they do not take a specific value, in complex models logical relationships (*Constraints*) will have to be defined among them.

According to this approach, a *Business Process Model* is defined as a set of *Allocation Types*. The relationship that starts with the solid rhomboid is a Composition Relationship. Therefore, in this layer, the proposed metamodel encompasses in a generic manner definitions of infrastructure allocation optimization problems.

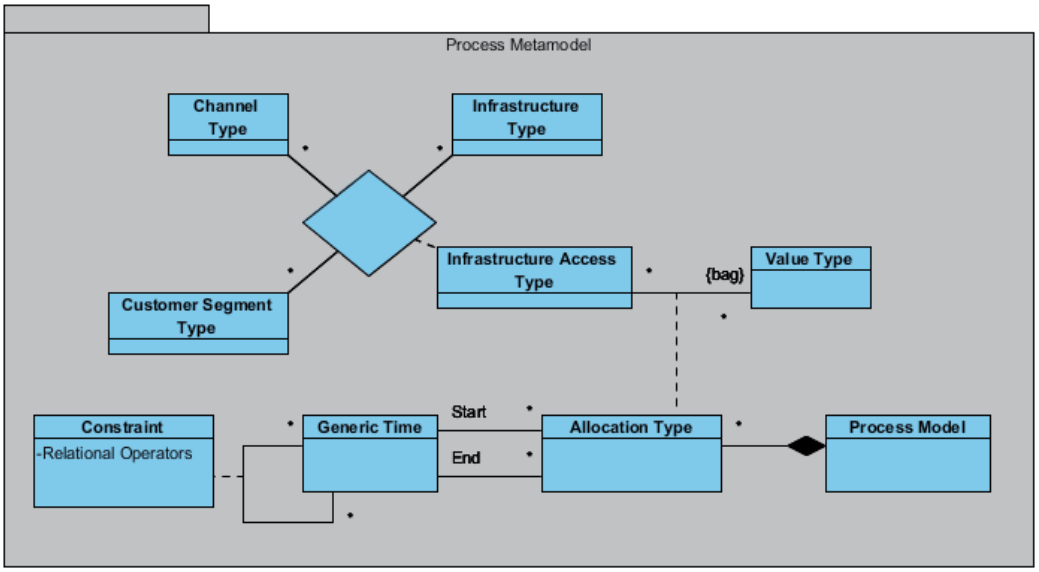


Fig. 2. UML class diagram of the Process Metamodel, the model of the process models

Model families can be derived from the instance patterns. Specifically, families can be defined if a dependency is found between two of the following entities: *Channel Type*, *Customer Segment Type* and *Infrastructure Type*; and according to which of these entities affects the definition of the prices or values, *Value Type*. As an example, there might be cases in which customers can access through various channels a given infrastructure (e.g.

reservation of a plane seat through the web or through a travel agency), and the combination of channel and infrastructure will define the types of value/ tariff that will be offered. In this case, the tariff would not depend on the customer type. Complementing the example, alternatively, the proposed model also encompasses another family of optimization models, such as when customers are segmented according to their loyalty to the company, and thus tariffs are in fact dependent on the customer segment.

4. Business process model

As stated in the hierarchy definition, the Business Process Model abstraction layer corresponds to the business process models generated through direct instantiation from the metamodel described in the former section.

We illustrate our proposal through a highly simplified model of a hotel room reservation process. In this simplified process we consider only one *Channel Type* (*CH1*) through which the customers place their reservations for a unique type of room (*R1*), which will be the only *Infrastructure Type* in this case. Customers are segmented in two types depending on whether they are only willing to make the reservation if they are offered a discounted price (*B*) or they are willing to pay the normal price (*A*; these customers would naturally also accept the discounted price). Therefore, there are two pre-specified *Value Types* (*Discounted* and *Full*) for each room. The business process designed by the hotel contemplates offering rooms initially at a specially discounted price and at some point in time, switching the price from *Discounted* to *Full*. The price will then remain *Full* for the rest of the time horizon. Table 1 shows the instances that would in this case form the business process according to the proposed model. *Generic Times* T1 and T3 define the start and the end of the process to be optimized. *Generic Time* T2 defines when the hotel switches from the *Discounted* price to the *Full* price.

Customer Segment Type	Channel Type	Infrastructure Type	Value Type	Generic Start	Generic End
A	CH1	R1	Discounted	T1	T2
A	CH1	R1	Full	T2	T3
B	CH1	R1	Discounted	T1	T2

Table 1. Example of room reservation business process model

5. Business process instance

The Business Process Instance is the third abstraction layer of the modeling hierarchy. This layer stems from assigning values to the optimization parameters linked to each model element. Optimization variables are signified by not having a value assigned in this layer. This approach allows establishing a taxonomy of infrastructure allocation problems, based on the set of parameters and variables that have been defined.

Taking the example of the former section, it is important to highlight that even such a simple business model lends itself to several optimization approaches. Specifically, it could lead to the following cases:

- Given the arrival rate for each customer type, the number of rooms to be offered, the full and the discounted price and assuming that there are no channel access restrictions, determine the optimal time T2 to maximize the hotel profit.

- Given the arrival rate for each customer type, the number of rooms to be offered, the time T2 at which the price changes and assuming that there are no channel access restrictions, determine the full and the discounted price that would maximize the hotel profit.

Both cases correspond to the same business process model. They are, therefore, instances of this process. Figure 3 depicts the UML model of process instances, in which, following the same approach as Gutierrez *et al*. (2006), each basic generic entity of the metamodel is associated to an instance that contains the parameter values and eventually the optimization variables. It is consequently at this layer that optimization problem definition takes place. There will be a group of optimization problems depending on the set of known parameters and on the set of variables to be optimized. Whereas the first two levels describe families of process models, in this level we find the organized collection of optimization problems that can be associated with those models.

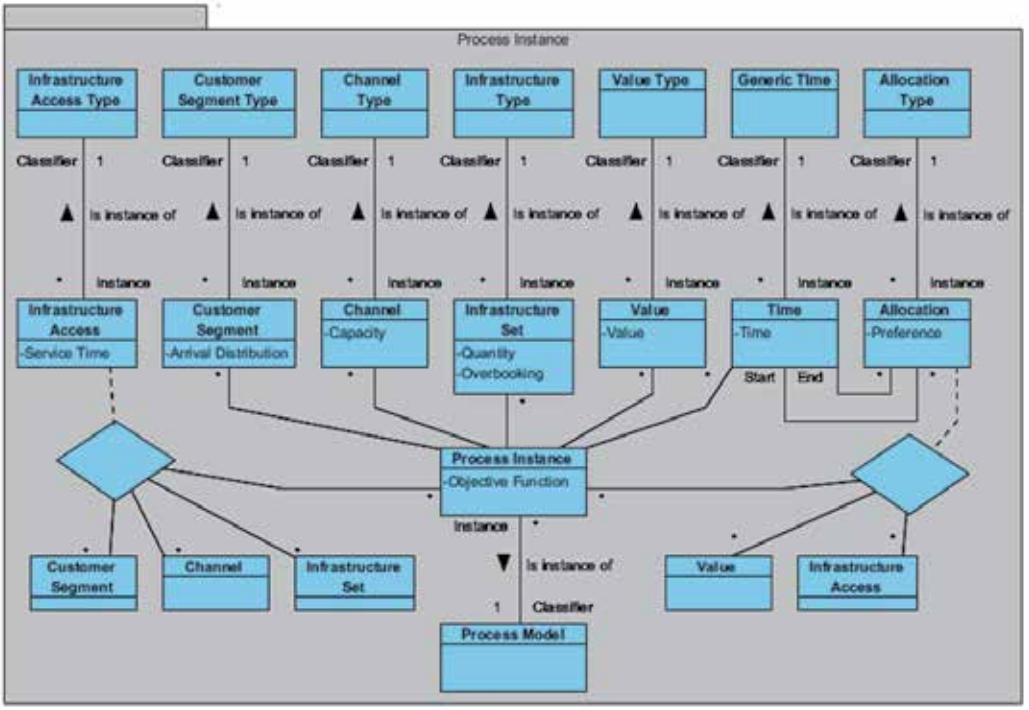


Fig. 3. UML class diagram of the Process Instance definition

We have included only a reduced number of representative parameters/variables for each entity as a means of illustration of the model capabilities. When trying to solve different infrastructure allocation problems, the necessity of further parameters will appear. Nevertheless, it is important to underline that the generic approach taken allows handling the natural extension of the group of parameters without extra coding or compilation (Gutiérrez *et al*. 2006). The user interface component of the DSS for which the model base is designed, should help the definition of entity attributes at this level. These attributes would be inserted in the corresponding database in an automatic pre-compiled procedure.

To enable the handling of problems in which a customer can be assigned to multiple intervals of time, it becomes necessary to make use of the attribute *Preference* of the entity *Allocation*. The attribute *Preference* will be used by the allocation algorithm to decide which slot to assign to a specific customer in case there are different alternatives. The DSS built upon the model base should ease the definition of a set of rules intended to define the values of the attribute.
Finally, the DSS should as well assist in the definition of an objective function, which is an attribute of the *Process Instance* entity, built upon mathematical combinations of a subset of parameters and variables.

6. Example of application to the health care sector

Based on the few cases reported of application of Revenue Management techniques to the Health Sector, and specifically on the pioneering work of Chapman & Carmel (1992) plus the above mentioned more recent work by Gupta & Denton (2008) and Nordgren (2009), an application to the dimensioning of hospital emergency wards is presented here.
Specialized doctors are treated as the infrastructures whose assignment is to be optimized (cardiologists, general surgeons, pediatricians...). Patients (system customers) are segmented according to their age (children vs. adults) and their symptoms. Channel Types to access the emergency ward are defined as being admitted though reception or through ambulance, distinguishing two levels of severity.

Customer Segment Type
ID_CustomerSegType
Adult: general discomfort
Adult: cardiac arrest
Adult: digestive problems
Adult: injuries
Child: sick

Channel Type
ID_ChannelType
Ambulance: slight
Ambulance: severe
Reception: slight
Reception: severe

Infrastructure Type
ID_InfrastructureType
Family Doctor
Cardiologist
Gastroenterologist
Traumatologist
Surgeon
Pediatrician

Value Type
ID_ValueType
High
Medium
Low

Generic Time
ID_GenericTime
T1
T2
T3

Table 2. Health Care example direct instances

The combined consideration of these three criteria leads to the evaluation of infrastructure access types as high, medium or low value from the point of view of the system's effectiveness. The Generic Times are used to define emergency shifts, since it might make sense not to have certain specialists in some shifts. In the example, three generic times represent two different shifts. Table 2 shows the example values for the model definition.
The process model is defined as the set of possible combinations of the direct instances of the process metamodel entities. Table 3 represents the business process example. Once the process model has been defined, several optimization problems can be stated in the business process instance level. In this example, we describe an infrastructure dimensioning problem, in which the number of medical specialists of each type is to be determined. Thus, in the process instance model of Figure 3, all the attributes of the instance entities but the Quantity attribute of the Infrastructure Set, will receive a value. The Quantity attribute will thus be the variable to be optimized. Attributes that do not have to be considered in the optimization problem—in this example O*verbooking* does not make sense— will receive a default value to avoid confusion with the variables of the problem.

Allocation Type					
ID_CustomerSegType	ID_ChannelType	ID_InfrastructureType	ID_ValueType	ID_GenTime_Start	ID_GenTime_End
Adult: general discomfort	Reception: severe	Family Doctor	High	T1	T3
Adult: general discomfort	Reception: slight	Family Doctor	Medium	T1	T3
Adult: cardiac arrest	Ambulance: severe	Cardiologist	High	T1	T3
Adult: cardiac arrest	Reception: severe	Cardiologist	High	T1	T3
Adult: cardiac arrest	Ambulance: severe	Family Doctor	Low	T1	T3
Adult: cardiac arrest	Reception: severe	Family Doctor	Low	T1	T2
Adult: cardiac arrest	Reception: severe	Family Doctor	Medium	T2	T3
Adult: digestive problems	Reception: slight	Gastroenterologist	Medium	T1	T2
Adult: digestive problems	Reception: slight	Gastroenterologist	Low	T2	T3
Adult: digestive problems	Reception: slight	Family Doctor	Medium	T1	T3
Adult: injuries	Ambulance: severe	Surgeon	High	T1	T3
Adult: injuries	Reception: severe	Surgeon	High	T1	T3
Adult: injuries	Ambulance: slight	Traumatologist	Medium	T1	T3
Adult: injuries	Reception: slight	Traumatologist	Medium	T1	T3
Child: sick	Ambulance: severe	Pediatrician	High	T1	T3
Child: sick	Ambulance: slight	Pediatrician	Medium	T1	T2
Child: sick	Ambulance: slight	Pediatrician	Low	T2	T3
Child: sick	Reception: severe	Pediatrician	High	T1	T3
Child: sick	Reception: slight	Pediatrician	Low	T1	T3
Child: sick	Reception: slight	Family Doctor	Medium	T1	T3

Table 3. Process model of the Health Care example

Infrastructure Access					
ID_ProcessModel	ID_ProcessInstance	ID_CustomerSegType	ID_ChannelType	ID_InfrastructureType	Service Time
P1	I1	Adult: general discomfort	Reception: severe	Family Doctor	1
P1	I1	Adult: general discomfort	Reception: slight	Family Doctor	0,5
P1	I1	Adult: cardiac arrest	Ambulance: severe	Cardiologist	1,5
P1	I1	Adult: cardiac arrest	Reception: severe	Cardiologist	2
P1	I1	Adult: cardiac arrest	Ambulance: severe	Family Doctor	2,5
P1	I1	Adult: cardiac arrest	Reception: severe	Family Doctor	3
P1	I1	Adult: digestive problems	Reception: slight	Gastroenterologist	0,5
P1	I1	Adult: digestive problems	Reception: slight	Family Doctor	0,5
P1	I1	Adult: injuries	Ambulance: severe	Surgeon	3,5
P1	I1	Adult: injuries	Reception: severe	Surgeon	4
P1	I1	Adult: injuries	Ambulance: slight	Traumatologist	0,5
P1	I1	Adult: injuries	Reception: slight	Traumatologist	0,5
P1	I1	Child: sick	Ambulance: severe	Pediatrician	1
P1	I1	Child: sick	Ambulance: slight	Pediatrician	0,5
P1	I1	Child: sick	Reception: severe	Pediatrician	1,5
P1	I1	Child: sick	Reception: slight	Pediatrician	0,5
P1	I1	Child: sick	Reception: slight	Family Doctor	1

Table 4. Infrastructure Access instances in the Health Care example

Infrastructure Set				
ID_ProcessModel	ID_ProcessInstance	ID_InfrastructureType	Quantity	Overbooking
P1	I1	Family Doctor	NULL	0
P1	I1	Cardiologist	NULL	0
P1	I1	Gastroenterologist	NULL	0
P1	I1	Traumatologist	NULL	0
P1	I1	Surgeon	NULL	0
P1	I1	Pediatrician	NULL	0

Table 5. Infrastructure Set instances in the Health Care example

Tables 4 and 5 show representative database record examples. A database implementation would require the utilization of proper numerical identifiers as primary keys of the main entities. Besides, it is necessary to include both the Process Model identifier and the Process Instance identifier in all the instance entities, in order to have multiple examples handled by a unique DSS.

7. Extensions and challenges

Hotel and Health Care examples illustrate the variety of models that can be defined within the proposed generic modeling hierarchy. In this final section we first discuss further possibilities of the model base as well as its limitations, and we identify the main promising extensions of the model; in summary, the challenges to be overcome in order to develop a generic DSS capable of coping with Revenue Management infrastructure allocation problems.

The model base, as it is defined in the previous sections, allows to model infrastructure allocation optimization problems in which different segments of customers, through a set of channels, can access a system to place a request for the use of an infrastructure during an interval of time. In the previous examples, demand for an infrastructure can only take place in a unique interval of time, i.e., in case no infrastructure is available in the defined interval, there is no attempt of allocation in another interval. Nevertheless, the model can handle more generic RM scheduling problems in which each customer segment might be offered the use of an infrastructure in different time slots. Even more, it would be possible to assign different values to the different slots. For instance, in the Health Care sector it makes sense to design a system in which there are some intervals which will be preferably assigned to some segments of customers.

In practice, applicability will be restricted to RM problems that do not require handling highly detailed time slots, since using the model base for such purposes would become tedious and they are better addressed by classic schedulers. On the other hand, the model does not support multi-slot allocation systems, such as medical procedures in which it is known that the first appointment will bring along ulterior ones.

The main extension of our proposal is to build a DSS upon the model base described throughout this chapter. In terms of the Sprague's (1980) classic framework, this would imply developing, on the one hand, an algorithmic module intended to solve the optimization problems, and on the other hand, a user interface module that should ease the definition of the allocation problems. In section 5 we mentioned the main functional requirements for the user interface module. With respect to the algorithmic module, solving

the RM problems can be accomplished by different techniques that may be classified in two main groups:

- Algorithmic optimization. Since the purpose of the model base is to cover a generality of infrastructure allocation problems, it is necessary to incorporate a generic algorithmic solution. Among the different optimization techniques, we find particularly interesting to explore the capabilities of Constraint Programming (CP). CP shows natural fitting with this generic approach due to three main characteristics (Tsang, 1993): in the first place, its flexibility in real-time definition of variables, constraints and objective functions; in the second place, the use of domain variables which makes it easier to handle infrastructure access constraints; and thirdly, the ease to define complex multi-criteria objective functions. Nevertheless, it appears to be a hard challenge to achieve a pre-compiled engine capable of dealing with the diversity of optimization problem families that can be modeled with the proposed model base.
- Simulation-based optimization. When stochastic parameters are involved—in RM problems, typically customer arrivals and service times, and occasionally others like customer behavior and choice preferences—simulation appears to be the natural approach. We believe that the model base is extensible to handle a generality of RM simulation models. The extension would stem from defining a 4th layer in the modeling hierarchy which would correspond to the execution level. This layer would correspond to the database model of the DSS and encompass all the execution data associated with a set of simulations of the system to be optimized.

8. Conclusions

A generic model base design for a DSS intended to deal with the problem of maximizing revenue in the allocation of service infrastructures offers great flexibility and allows overcoming the intrinsic dependency of the algorithm on the business process model. This dependency hinders the application of RM techniques through traditional DSS. The hierarchical metamodel-model-instance approach proposed to define the RM DSS model base here shows great potential and addresses the three core elements of the mentioned algorithm dependency: The specific traits of the business processes, the decision variable(s) to be optimized and the business objective(s) to be achieved. It also supports the structured definition of business process typologies along these dimensions, which can lead to applications of generalized interest to the RM community regarding the definition of taxonomies of RM optimization problems. Furthermore, this approach allows defining and including in the DSS the specific parameters required in a given model without ad-hoc coding or compilation. Examples taken from the Hotel and Health Care sectors illustrate the potential of the proposed generic approach. The modeling hierarchy can be extended to encompass the data model of the RM DSS database component.

9. Acknowledgment

This work is supported by the research project DPI2008-04872, “Optimization of service infrastructures assignment through simulation - hotel and health sectors” (Spanish National Research Plan) and by the Human Resources Mobility Action I-D+i 2008-2011 of the Spanish Ministry of Education for a visiting scholar stay at MIT Center for Transportation & Logistics.

10. References

Chiang, W.C; Chen, J.C.H. & Xu, X. (2007). An overview of research on revenue management: Current issues and future research, *Int. J. Revenue Management*, Vol.1, No.1, (January 2007), pp. 97-128, ISSN 1474-7332

Eriksson, H.-E. & Penker, M. (2000). *Business Modeling with UML: Business Patterns at Work*, John Wiley & Sons, ISBN 978-0471295518, NY, USA

Fowler, M. (2004). *UML Distilled: A Brief Guide to the Standard Object Modeling Language*, Addison-Wesley, 3rd ed., Object Technology Series, ISBN 978-0321193681, Boston, USA

Gupta, D. & Denton, B. (2008). Appointment scheduling in Health Care: Challenges and opportunities, IIE Transactions, Vol.40, No.9 (Sep. 2008), pp. 800-819, ISSN:0740-817X

Gutiérrez, M.; Durán, A. & Cocho, P. (2006). An alternative approach to the standard Enterprise Resource Planning life cycle: Enterprise reference metamodeling, *Lecture Notes in Computer Science,* Vol.3982, (May 2006), pp. 964-973, ISSN 0302-9743

Muhanna, W.A. & Pick, R.A. (1994). Meta-modeling concepts and tools for model management: A systems approach, *Management Science*, Vol.40, No.9 (Sep. 1994), pp. 1093-1123, ISSN: 0025-1909

Nordgren, L. (2009). Value creation in health care services: Developing service productivity, *International Journal of Public Sector Management*, Vol.22, No.2 (March 2009), pp.114-127, ISSN 0951-3558

OMG (May 2010). OMG Unified Modeling Language (OMG UML), Infrastructure Version 2.3, Available from:
http://www.omg.org/spec/UML/2.3/Infrastructure/PDF/

OMG (Feb 2011). Introduction to OMG's Unified Modeling Language™ (UML®), Available from : http://www.omg.org/gettingstarted/what_is_uml.htm

Power, D.J. (2002). *Decision Support Systems : Concepts and Resources for Managers,* Greenwood, ISBN 1-56720-497-X, Westport, CT, USA

Sprague, R.H. (1980). A framework for the development of Decision Support Systems, *MIS Quaterly,* Vol.4, No.4, (December 1980), pp. 1-26, ISSN 0276-7783

Talluri, K.T. & Van Ryzin, G.J. (2004). *The Theory and Practice of Revenue Management,* Springer, ISBN 1-4020-7701-7, NY, USA

Tsang, E. (1993). Foundation of Constraint Satisfaction, Academic Press, ISBN 0-12-701610-4, London, GB

Zaki, H. (2000). Forecasting for airline revenue management, *The Journal of Business Forecasting Methods & Systems,* Vol.19, No.1, (January 2000), pp. 2-6, ISSN 0278-6087

11

An Efficient Failure-Based Maintenance Decision Support System for Smalland Medium Industries

M. A. Burhanuddin[1], Sami M. Halawani[2] and A.R. Ahmad[2]
[1]*Universiti Teknikal Malaysia Melaka,*
[2]*King Abdulaziz University Rabigh*
[1]*Malaysia*
[2]*Kingdom of Saudi Arabia*

1. Introduction

Small and Medium Industries (SMI) are the key contributors to economic growth in developing countries. SMI contributes in generating employment and engage in small and medium-scale manufacturing lines to generate profit and employment.

In fact, the largest portion of manufacturing firms fall into SMI categories and the SMI businesses are the backbone of the large-scale industry (Shamsuddin et al. (2004)). However, Junaidah (2007) reported that SMIs lacking appropriate decision making. According to her, SMI do not have a good system to evaluate the overall system of machines, contractors performance and their business's principles. Therefore, maintenance decision support system (DSS) is essential to ensure maintainability and reliability of equipments in industries. Poor machinery maintenance management will result in capacity loss, poor product quality and customer dissatisfaction. These downturns usually occur depending on the efficiency of the reliability programs executed by the organization.

This chapter reveals on the important DSS models i.e. Decision Making Grid (DMG) to be embedded with computerized maintenance management system (CMMS) to aid maintenance strategies for the machines as an adoption of technology management in SMI. Next this chapter demonstrates on how DMG model can be used as a decision support module in Failure-based Maintenance (FBM).

2. Brief description of the maintenance issues in SMI

Shamsuddin et al. (2004) conducted a survey to study FBM issues faced by SMI. They have listed issues related to equipment maintenance, as follows:

i. Lack of human resources, both in terms of number and skill or expertise;
ii. Emphasis on short-term gains and lack of long-term plans;
iii. Lack of state-of-the-art modern technology;
iv. Lack of understanding about the role of technology;

v. Insufficient funding for machinery investment;
vi. Lack of time to think, and reengineering is expensive;
vii. Operators have poor technical knowledge about the machine they are operating;
viii. Poor participation from non-manufacturing units such as administration, marketing, and purchasing i.e. looking at the system from the point of sub optimization, which is contrary to Total Productive Maintenance or Total Quality Maintenance practices;
ix. Overall low level equipment effectiveness evaluation, especially on availability, performance rates, and quality rates; and
x. Slow response of the contractors on maintenance work.

Saleh and Ndubisi (2006) highlighted that SMI lack a comprehensive framework to be used in solving critical issues. Later, Kittipong (2008) conducted more comprehensive studies on technology relationships in SMI. He conducted the surveys and identified 20 factors of technology relationship and innovations. He conducted hypotheses and concluded that technology expertise is the first priority for SMI success in the area of manufacturing. SMI should start to look at the technology escalation procedure and manage their vendors or contractors efficiently. He has suggested a further research direction for factorial and decisional analyses using longitudinal data in his thesis. This motivates this study to aid maintenance decision-making in FBM for SMI.

3. Effects of the machinery failures in SMI

The primary functions of most machines in industries are concerned, in some way, with the need to earn revenue or to support revenue earning activities. Poor machinery maintenance will lead to more emergency breakdowns. The breakdowns affect the production capability of physical assets by reducing output, increasing operational costs, and, thus, interfering with customer services. Machinery failures in SMI production lines will increase the operation cost and reduce their profit margin.

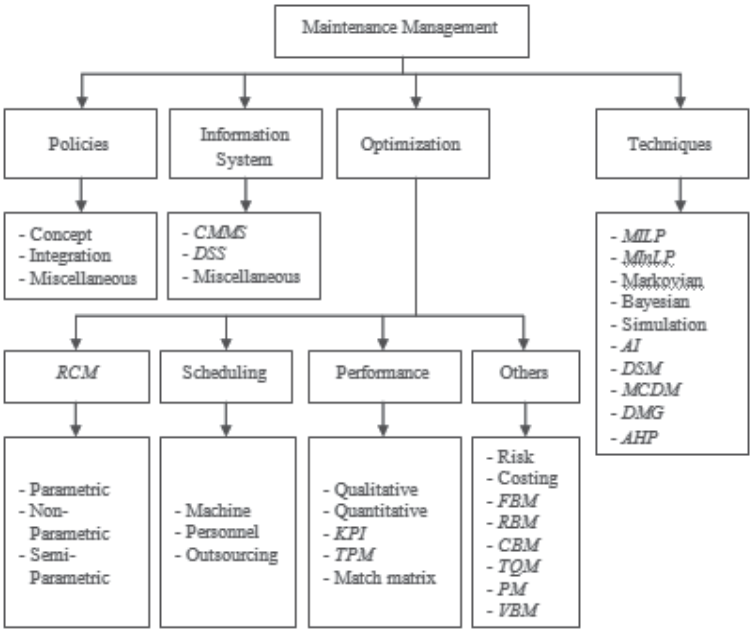
In brief, the effects of downtime are much greater than the cost of repairing the failures. For example, if a filling machine fails in a SMI production line, the end products will spill over. This also results in labour safety issues as well as business losses.

4. Related works

It has been a long journey for the evolution of maintenance management models and techniques from 1940 until now. Maintenance is a wide area, and involves planning, coordinating, controlling, supervising and managing activities in order to achieve the organization's purposes and goals. Palmer (1999) defines maintenance management as control that must be taken to keep the equipment in its working state, by preserving it from any failures. The goals of the maintenance study are to reduce downtime and cost, while improving the adequacy and quality of service to maximize the availability of equipment in an operative state. There are many literatures available from various researchers and practitioners in the field of maintenance management.

In general, Burhanuddin (2009) narrowed down maintenance literature into four main-classes, sub-classes and sub-divisions, as shown in Figure 1.

Some observations of the study in each of the above four main-classes in maintenance management and their fractions are presented in detail as follows.



Abbreviation represent: CMMS: Computerized Maintenance Management System; DSS: Decision Support System; RCM: Reliability-centred Maintenance; KPI: Key Performance Index; TPM: Total Productive Maintenance; FBM: Failure-based Maintenance; RBM: Risk-based Maintenance; CBM: Condition-based Maintenance; TQM: Total Quality Maintenance; PM: Preventive Maintenance; VBM: Vibration-based Maintenance; MILP: Mixed Integer Linear Programming; MInLP: Mixed Integer non-Linear Programming; AI: Artificial Intelligence; DSM: Decision Support Model; MCDM: Multiple Criteria Decision-Making; DMG: Decision-Making Grid; and AHP: Analytical Hierarchical Process

Fig. 1. Sub-division Tree of Maintenance Management

4.1 Maintenance policies

Guy and Steve (1999) reviewed the general terms of maintenance. They studied maintenance development and the reason behind it changing from time to time. After that, they focused on the enhancement and evolution of information systems in maintenance.
Some important findings on enhancement of the development of information systems in maintenance are given, as they have discovered that new strategic development in maintenance is due to information system application and development.
Lam and Lin (2004) integrated some replacement policies into corrective maintenance. Corrective or Failure-based Maintenance (FBM) is unscheduled maintenance or repair to return the machine to a defined state. There are no interventions until a failure has occurred. Lewis (1999) addressed corrective maintenance as reactive maintenance, where any emergency breakdown will lead to a bigger impact on the operation. Since the failures are unplanned, they might result in a big loss to the organization in terms of cost and time. Therefore, a better maintenance concept must be introduced to prevent this unplanned downtime and reduce the cost of the failure. The breakdown maintenance concept is still applied to equipment that is not mission critical and where the downtime would not affect the main operation of the organization, such as light bulbs and consumable parts. Lam and Lin (2004) have introduced some numerical methods and designed optimal replacement policies in FBM.

According to Palmer (1999), if the plant spends a lot of time on breakdown maintenance, then it does not spend enough time on preventive maintenance. Thus, predictive maintenance is introduced to resolve deficiencies with scheduled downtime. Predictive maintenance involves monitoring certain conditions or variables associated with the equipment (Bentley, 1993). The simplest method of condition-based monitoring involves the four human senses of sight, sound, touch and smell to predict a failure. Repairing activities take place when the condition shows that a failure may be imminent.

Unlike the Condition-based Maintenance (CBM) policy, in predictive maintenance the acquired controlled parameters data are analyzed to find a possible temporal trend. This makes it possible to predict when the controlled quantity value will reach or exceed the threshold values. The maintenance staffs will then be able to plan when, depending on the operating conditions, the component substitution or revision is really unavoidable. The following main activities can help to avoid machine breakdown (Bentley, 1993):

i. Fault detection,
ii. Fault isolation,
iii. Fault elimination, and
iv. Verification.

The same was demonstrated by Geert and Liliane (2004), who distinguished five basic maintenance policies in corrective and preventive maintenance. Preventive maintenance is carried out using a planned, periodic and specific schedule to keep a machine in a stated working condition, throughout the process of checking and reconditioning. Preventive maintenance is defined as the pre-breakdown model performed on equipment to either eliminate or reduce the emergency failures within pre-determined economic limits (Lewis, 1999). The model has been introduced to enhance reliability and confidence of the machine in advance. This maintenance is categorized as proactive maintenance. The service is repeated at a pre-determined frequency to avoid any unplanned breakdown of the machine. Bentley (1993) divides the preventive maintenance (PM) model into CBM with two main categories, as follows:

i. Monitored PM, which involves monitoring when the machine is in operation. Triggers on any potential failure will be detected. Repair activities are to be conducted before any unplanned breakdown; and
ii. Scheduled maintenance, where the service is being conducted on the same machine at specific counter or time intervals. The maintenance crew always follows the standard checklist to conduct PM activities, which involves scheduled replacement of parts, service, alignment, greasing, lubrication, confidence testing, etc.

Geert and Liliane (2004) distinguished five basic policies in maintenance, as follows:

i. Failure-based Maintenance (FBM);
ii. Use-based Maintenance (UBM);
iii. Condition-based Maintenance (CBM);
iv. Detection-based Maintenance (DBM); and
v. Design-out Maintenance (DOM).

They have suggested seven steps to follow in a modular framework on maintenance policies, before building any maintenance policies, as follows:

i. Identification of the objectives and resources,
ii. Selection of the most important maintenance systems,
iii. Identification of the most critical machines and their components,

iv. Maintenance policy selection,
v. Optimization of the maintenance policy parameters,
vi. Implementation and evaluation, and
vii. Feedback.

Later, Geert and Liliane (2007) discovered more findings. They have designed concrete policies and developed a framework to set up an industrial management centre. There is a group of people in the centre to control all maintenance work and escalate the work to the respective site. Their model can be used to develop a customized maintenance concept that is suitable for multinational companies.

Since centralized maintenance control in a centre requires more people, Dhillon and Liu (2006) presented the impact of human errors on maintenance. They conducted a literature survey of human errors in maintenance and their impact on the manufacturing plant. They reported that human errors will reduce production profit. Imad (2007) highlighted the role of maintenance as a profit-generating functionality, by introducing a maintenance quality concept in the manufacturing system. He has intimated the relationship between a manufacturing system's capacity and the total manufacturing costs per unit of quality item, in order to describe and illustrate how maintenance generates profit to the company.

Finally, Imad (2007) proved that maintenance is not a cost centre, but a profit-generating centre in the manufacturing sectors. Attempting to fulfil maintenance as a profit-making centre, Alexandre et al. (2008) introduced the excellent-maintenance concept of doing more work with fewer people and less money. They stressed this new generation maintenance concept, which includes:

i. Remote maintenance: By leveraging information, wireless and internet technologies, users may log in from anywhere and with any kind of device as soon as they get an internet connection. The maintenance operation team can connect remotely to the equipment in the factory, then run setup, control, configure, diagnose, debug, fix, monitor performance, or download data for analysis;
ii. Cooperative or collaborative maintenance: Excellent-maintenance symbolizes the opportunity to implement an information infrastructure connecting geographically dispersed subsystems;
iii. Immediate or on-line maintenance: The real-time remote monitoring of equipment status, coupled with programmable alerts, enables the maintenance operator to respond whenever any breakdown occurs. In addition, high-rate communications allow them to quickly obtain several kinds of expertise and to accelerate the feedback reaction in the local loop, connecting the product, monitoring agent and maintenance support system together. It has almost unlimited potential to reduce the complexity of traditional maintenance guidance through on-line guidance, based on the results of decision-making and analysis of product condition;
iv. Fault diagnosis or localization: Excellent-maintenance offers experts the ability to perform on-line fault diagnosis, share their valuable experiences with each other, and suggest remedies to the operators if an anomalous condition occurs in the inspected machine. In addition, lock-outs and isolation can be performed and recorded on location. Consequently, the amount of time it takes to communicate a production problem to the potential expert solution provider can be reduced, the quality of the information shared can be improved, and thereby the resolution time reduced. All these factors increase the availability of production and facilities equipment, reduce the mean time taken to repair, and significantly utilize field service resources and costs;

v. Repair or rebuilding: Remote operators could, via an electronic connection, tap into specialized expertise rapidly without any travelling or scheduling delays. Downtimes could be conceivably reduced through direct troubleshooting with source designers and engineers. In addition, diagnosis, maintenance work performed, and parts replaced are documented on the spot, through structured responses to work steps displayed on the mobile workstation; and
vi. Modification or improvement: The multi-source knowledge and data environment provided by excellent-maintenance allows efficient information sharing. With the availability of tools for interacting, handling, and analysing information about product state, the development of maintenance engineering for product lifecycle support, including maintenance and retirement stages such as disassembly, recycling, reuse and disposal, is becoming feasible.

The excellent-maintenance concept requires a good information technology system. There are very promising developments in information technology, which can help to improve maintenance practice. However, the information system development in maintenance is still relatively young in the business area, and its recent development is discussed in the next section.

4.2 Maintenance information system

The information system is becoming an important tool for achieving efficiency and effectiveness within maintenance, provided that the correct and relevant information technology is applied. In fact, there are many papers on the maintenance information system and technologies, where they are always integrated with other disciplines such as inventory control, supply chain management and communication technology on the manufacturing shop floor. Here, four reviews are worth discussing under this sub-section. The first review is given by Labib (2004), who discusses the CMMS development to facilitate the management of maintenance resources, to monitor maintenance efficiency and to provide appropriately analysed management information for further consideration. Labib (2004) has suggested some benefits that CMMS can offer, as follows:

i. It allows operators to report faults faster by filling-up the electronic-form. At the same time, it enables the maintenance team to respond to and update the form promptly;
ii. It can facilitate improvements in communication between operators and maintenance personnel, and is influential in ameliorating the consistency of information flow between these two departments;
iii. It offers insight into wear-and-tear in FBM activities;
iv. It provides maintenance planners with the historical information necessary for developing next PM schedules; and
v. It can track the movement of spare parts and requisition replacements whenever necessary.

In the second review, O'Donoghue and Prendergast (2004) proved that CMMS is very beneficial. They examined the basis of various maintenance management strategies used to date in international manufacturing. They have demonstrated how CMMS is used to capture maintenance activities, and analysed maintenance time and cost in an Irish textile manufacturing company.

Lately, rapid growth in computer technology has opened up new possibilities in CMMS development. This brings us to the third review, in which Mirka (2008) reported that CMMS

provides tremendous benefit to maintenance. She identified salient characteristics of CMMS by showing that information technology investment has a positive correlation to company profitability and competitiveness. She also developed information technology tools based on a company's factors of goal and purpose.

Despite providing significant characteristics of CMMS, which may fit in with the needs of the industries, Labib (2004) also discovered that the majority of available CMMS in the market lack decision support for management. Last but not least, the fourth crucial review, highlighted by Sherif et al. (2008). They managed to embed DSS into CMMS as an advanced software engineering approach. DSS constitutes a class of computer-based information systems, including knowledge-based systems that support decision-making activities. A computer program is written using available maintenance techniques to automate the analysis and results. The program is executed in sub-procedures in CMMS as a DSS module. Sherif et al. (2008) synthesized DSS with problem-solving in concrete bridge decks maintenance activities. They proposed different decisions for different types of repair, i.e. shallow repair, deep repair, protective repair, non-protective repair, and deck replacement. All decision-makers must consider the cost of repair when making any recommendation. Sherif et al. (2008) also deliberated on how DSS is used to model human reasoning and the decision-making process for concrete bridge decks. At the end of the study, they concluded that buying sophisticated hardware or software is not the complete answer. However, justification on middleware software, and an object-oriented system by integrating some maintenance techniques into the DSS, is another potential area to consider. There are various available techniques in maintenance that can be programmed into CMMS and can measure maintenance activities. The techniques are elaborated on in the next section.

4.3 Decision support system and optimization techniques

DSS gathers and presents data from a wide range of sources in a way that can be interpreted by humans. Important features of DSS are given as follows (Williams and Sawyer, 2006):

i. lnputs and outputs;
ii. Assist tactical-level managers in making tactical decisions; and
iii. Produce analytical models, such as mathematical representation of a real system.

A quantitative approach in the DSS model allows maintenance manager to play a simulation what-if game to reach decisions. They can simulate certain aspect of the organization's environment in order to decide how to react to a change in the conditions affecting it. By changing the hypothetical inputs to the maximum and minimum levels, the managers can see how the model's outputs are affected. There are four aspects to maintenance optimization models, as follows (Amik and Deshmukh (2006)):

i. Description of a technical system, its function and importance;
ii. Modeling of the deterioration of the system in time and possible consequences for this system,
iii. Description of the available information about the system and action open to management and
iv. Objective function and an optimization technique, which helps in finding the best practice.

An efficient DSS for SMI should have some real time data analysis capabilities, and be able to query from the database, perform the calculation and provide decisions such as:

i. Which machines should go for FBM, fixed time maintenance, design-out maintenance, condition-based maintenance, preventive maintenance, etc.

ii. When the next maintenance is due for every machine;
iii. Which contractor should be called to perform the maintenance work;
iv. Estimates of man-hours required;
v. Description of the tasks involved and how much time is required;
vi. Lists of all required replacement parts and their locations;
vii. Forecast on the spare parts, tools and their costs;
viii. Re-order level of the machine parts and other accessories; and
ix. An estimate of the maintenance priorities and their impact.

The most prominent objective of the various techniques in maintenance DSS is to supply vital data and evidence, to derive better strategies to minimize machine downtime and maintenance cost. DMG is one of the famous techniques in DSS as suggested as follows.

5. Decision-making grid

The maintenance Decision-Making Grid (DMG), introduced by Labib (1998) acts as a map where the performances of the worst machines are placed based on multiple criterions, i.e. frequency of failures and downtime. The results provide guidelines for the action, which will lead to the movement of machines towards an improvement of the maintenance strategies with respect to multiple criterions. Labib (2004) defined the main input from the failures for DMG analysis as follows:

i. The response time;
ii. The diagnostic time;
iii. The repair time; and
iv. Frequency of failures.

Based on the input, machines are mapped into a two-dimensional matrix and appropriate maintenance strategies will then be implemented, such as total productive maintenance, reliability-centred maintenance, design-out maintenance, condition-based maintenance, fixed-time maintenance, etc.

There are many researchers who have studied the DMG and apply it in the equipment management area. Among those, there are three selected reviews that are worth discussing under this sub-section. In the first Labib (1998a) has introduced the DMG model to help maintenance management identify breakdown maintenance strategies. In short DMG is a control chart in two dimensional matrix forms. The columns of the matrix show the three criterions of the downtime, whilst the rows of the matrix show another three criterions of the frequency of the failures. The model consists of these three steps:

i. Criteria analysis;
ii. Decision mapping and
iii. Decision support.

Here, a better maintenance model for quality management can be formed by handling both the rows and columns of the matrix respectively. The matrix offers an opportunity to decide what maintenance strategies are needed for decision-making such as to practice Operate to Failure (OTF), Fixed Time Maintenance (FTM), Service Level Upgrade (SLU), Condition-based Maintenance (CBM), Designed-out Maintenance (DOM), Total Productive Maintenance (TPM) and Reliability-centered Maintenance (RCM).

The second important review was undertaken by Fernandez et al. (2003), in which implementation of DMG in CMMS was discussed in detail. They extended the theory of the maintenance maturity grid and implemented it into a disk brake pad manufacturing

company in England. The results can provide maintenance policies in the respective functional group in production lines, to achieve their common goal to reduce downtime.
Later, Zulkifli et al. (2008), in the third review, comprehended the model and demonstrated the hybrid intelligent approach using the DMG and fuzzy rule-based techniques. In their study, the DMG is employed in small and medium food processing companies to identify their maintenance strategies. DMG is used in these study as the model is flexible, and considers OTF, FTM, SLU, CBM, DOM, TPM and RCM strategies in the same grid. The model is able to analyze multiple criteria and is the best choice when the number of machines is less than fifty (Pascual et al, 2009). It can be used to detect the top ten problematic machines on the production floor with several system conditions. This is with regards to failures such as fatigue, imbalance, misalignment loosened assemblies, and turbulence, which can occur in rotational or reciprocating parts such as bearings, gearboxes, shafts, pumps, motors and engines. Identifying the top ten problematic machines is in alignment with the 80-20 rule. The rule states that eighty percent of the problems arise from the same twenty percent of the root causes.
ln another word, once twenty percent of the root cause had been fixed, then eighty percent of the problem is resolved. The application of the model can have a breakthrough performance, as it fulfils the purpose of the model to map machines into a certain grid in a matrix and suggests the appropriate maintenance strategies to comply with.

5.1 Development of the decision-making grid model

There are many publications on CMMS and DMG applications in the area of maintenance. Among them, there are three journal papers that are most related to this study of the DMG model, written by Labib (1998b and 2009) and Fernandez et al. (2003). Labib (1998b) introduced DMG as a maintenance decision-making tool to be embedded in CMMS. Later, Fernandez et al. (2003) included DMG as a sub-module in their CMMS. They tested the model in one of the brake pad manufacturing companies in England. Next Zulkifli et at. (2010) integrated DMG with a fuzzy rule-based hybrid approach. The DMG model was formed as a control chart by itself in two-dimensional matrix forms. The columns of the matrix show the three criterions of the downtime, whilst the rows of the matrix show another the criterions of the frequency of the failure.
A better maintenance model for quality management can be formed by handling both the rows and columns of the matrix respectively. The matrix offers an opportunity to decide what maintenance strategies are needed for decision-making such as to practice OTF, FTM, SLU, CBM or DOM. The matrix also suggests maintenance concepts that are useful for each defined cell of the matrix such as TPM or RCM approaches. The results can provide maintenance policies in the respective functional group in production lines to achieve their common goal, to reduce downtime. There are two basic steps to follow in the DMG model as follows:

Step 1. Criteria Analysis: Establish a pareto analysis of two important criteria:

a. Downtime, which is the main activity conducted by a maintenance crew; and
b. Frequency of breakdown calls, which is always a concern for a customer service centre.

   The objective of this phase is to assess how bad the worst performing machines are over a certain period of time. The machines, as regards each criterion are sorted and placed into a top ten list of high, medium, and low boundaries, which are divided into three categories using the tri-quadrant approach as follows (Burhanuddin (2009)). let x be the frequency of failures.

Let k = $\frac{x_{\max} - x_{\min}}{3}$, where max = maximum and min = minimum, then the intervals are obtained as:

Highfrequency $= [\, x_{\max}, x_{\max} - k \,)$ (1)

Mediumfrequency $= [\, x_{\max} - k, x_{\max} - 2k \,)$ (2)

Mediumfrequency $= [\, x_{\max} - 2k, x_{\min} \,]$ (3)

Similarly, let y be the downtime of the machines. Let l = $\frac{y_{\max} - y_{\min}}{3}$, then the intervals are obtained as:

Highdowntime $= [\, y_{\max}, y_{\max} - l \,)$ (4)

Mediumdowntime $= [\, y_{\max} - l, y_{\max} - 2l \,)$ (5)

Mediumdowntime $= [\, y_{\max} - 2l, y_{\min} \,]$ (6)

Using the above formulae, machines are mapped into a two-dimensional matrix based on their respective intervals on frequency of failures and downtime. Next, decision strategies for machines such as OTF, FTM, SLU, CBM, DOM, RCM or TPM can be implemented (Labib (1998b)).

Step 2. Decision Mapping: Those machines that meet both criteria and are ranked in Step 1, are then mapped on the grid as shown in Table 1.

Frequency	**Downtime**		
	Low	**Medium**	**High**
Low	OTF	*FTM1*	*CBM*
Medium	*FTM2*	FTM3	*FTM4*
High	*SLU*	*FTM5*	*DOM*

Table 1. Decision-Making Grid (Labib, 1998b)

In this phase the decision is developed by comparing two-dimensional matrices. The objective of this exercise is to implement appropriate actions that will lead to the movement of machines towards improved machine maintenance strategies, complying with Labib (1998b) in respect of the multiple criterions as follows:

a. Operate to failure (OTF): Machines that are very seldom failed (low frequency). Once failed, the downtime is short (low downtime). The machines in this region are performing well and are easy to repair
b. Skill levels upgrade (SLU): Machines that always fail (high frequency), but can be fixed quickly (low downtime). Basically, the problems faced by the machines in this grid are relatively easy to fix. Upgrading operators' skill levels in fixing the problem will help to reduce response time;
c. Condition-based maintenance (CBM): Machines that seldom fail (low frequency). However, this is a killer machine, as it takes a long time to bring it back to normal

operation (high downtime). Machines in this region should be monitored closely. It is a good idea to equip the machines with sensors or any condition-based monitoring devices to observe their performance;

d. Design-out maintenance (DOM): Machines that always fail (high frequency). Once failed, it takes a long time to bring them back to normal operation (high downtime). This region consists of the worst production machines and it is recommended that they be structurally modified. A maintenance department should consider a major design-out project and spend money to upgrade the machines in this grid; and

e. Fixed time maintenance (FIM): There are five categories as follows:
   - FTMI: Concern about the timing of the maintenance. Maintenance department should revise when is the best time to conduct the preventive maintenance;
   - FTM2: Maintenance department should study who is the most suitable person to conduct the maintenance job for those machines in this grid, i.e. operator, technicians, contractors or maintenance engineers;
   - FTM3: Failure frequency and downtime are almost at the moderate cases,
   - FTM4: Maintenance department should revise on their maintenance checklist and instructions. What are the instructions that must be included during maintenance; and
   - FTM5: Maintenance department should revise on how the maintenance should be conducted and improved.

The TPM approach should be applied for the lower triangle of the DMG matrix, as shown in Table 2 (Labib, 1998b). TPM is applied globally and one of the TPM concepts is to empower the operators to maintain continuous production on totally efficient lines (Nakajima,1988). TPM is the continuous knowledge transfer to operators and the maintaining of the production equipment together with the maintenance crew. Hence, we can slowly reduce the waiting time for technicians to be in the production plant. On the other hand it gives operator the opportunity to eliminate the root causes of the machines' errors on a small level, before they become big ones.

	Downtime		
Frequency	**Low**	**Medium**	**High**
Low	*OTF*	*FTM1*	*CBM*
Medium	*TPM*	*FTM3*	*FTM4*
High			*DOM*

Table 2. DMG-TPM Strategy (labib 1998b)

The RCM approach should be applied for the upper triangle of the matrix as shown in Table 3. RCM involved a study and measurement of the probability that a machine will operate as expected at the desired level, for a specific period of time, under the design operating conditions, and without any failures. Once the problematic machines are identified, the maintenance strategy should be adjusted to ensure the longest survival of the machine to complete a mission at a specific

time (Elsayed, 1996). Strategies such as CBM or DOM are executed based on measurements and estimates.

Once the top ten worst performing machines are identified from DMG, the appropriate action to identify the next focused action is implemented. In other word, maintenance staffs need to move from the strategic systems level to the operational component level.

Frequency	Downtime		
	Low	Medium	High
Low			
Medium	FTM2		
High	SLU	FTM5	

RCM

Table 3. DMG-RCM Strategy (Labib, 1998b)

5.2 Empirical results

As a case study, the dataset from Fernandez et al. (2003) on the disk brake pad manufacturing company in England is used. Table 4 shows the level of the frequency of failures decision analysis using formulas in (1), (2) and (3).

ID	Frequency	Level
A	23	High
B	19	High
C	16	Medium
D	16	Medium
E	16	Medium
F	16	Medium
G	15	Low
H	14	Low
I	13	Low
J	12	Low

Table 4. Frequency of Failure Decision Analysis (Fernandez et al., 2003)

Decision analysis graph on frequency of failures is shown in Figure 2.

Once intervals of the machine are categorized correctly using formulae (4), (5) and (6), the DMG matrix can be constructed as shown in Table 6. Next, a strategic decision is identified by mapping at DMG in Table 1, given by Labib (1998b). Then, a valid recommendation can be given to the SMI maintenance management team to implement maintenance strategy.

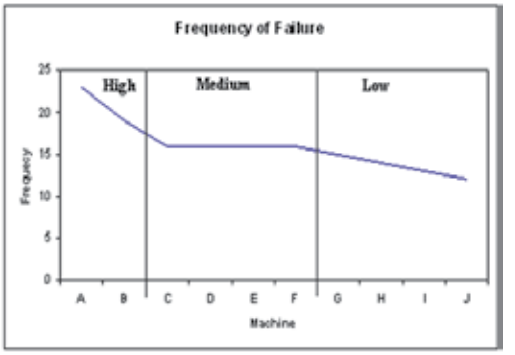


Fig. 2. Frequency of Failure Decision Analysis (Fernandez et al., 2003)

Table 5 shows the level of decision analysis on downtime in hours, from Fernandez et al. (2003) data using formulas in (4), (5) and (6).

ID	Downtime	Level
B	64	High
A	32	High
F	29	High
K	28	Low
H	27	Low
L	25	Low
C	25	Low
G	22	Low
D	20	Low
E	20	Low

Table 5. Decision Analysis on Downtime (Fernandez et al., 2003)

Decision analysis graph on downtime is shown in Figure 3.

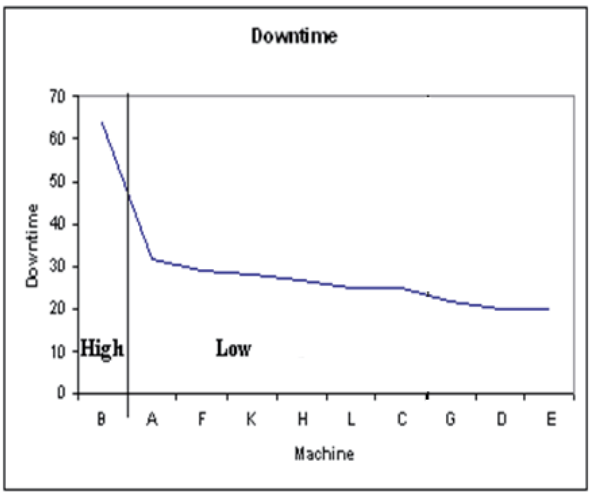


Fig. 3. Decision Analysis on Downtime (Fernandez et al., 2003)

7. Conclusion and further research

This chapter aim to improve maintenance strategies in FBM by extending some theoretical method introduced by Labib (1998b) and Burhanuddin (2009) in DMG model to provide maintenance strategies for SMI. The analysis provides insight by using collective approaches in two-dimensional matrices, and simultaneously measures multiple criterion of the downtime and frequency of failures.

Maintenance strategies are identified using the DMG model, based on important factors, including the machines' downtimes and their frequency of failures. The machines are categorized into the downtime criterions and frequency of failures, which are high, medium and low using tri-quadrant formulae. The experimental studies are conducted using maintenance dataset given by Fernandez et al. (2003). The proposed models can be used by decision makers to identify maintenance strategies and enhance competitiveness among contractors in FBM. There have been very promising development in computer science, which can help SMI to monitor and compare maintenance activities.

The analysis provides insight by using collative approaches in two-dimensional matrices, and simultaneously investigates three criterion of the downtime and frequency of failures. Tri-quadrant analysis is used to ascertain the model through the incorporation of multiple criterions of the decision-making analysis for SMI. The result of the analysis provides some decision-making strategies for machinery maintenance.

This study focuses on DMG based on few factors respond time, diagnostic time, repair time and frequency of failures. More research on organization procedures and underlying characteristics of the machines, such as their model, age, made, price, etc., can lead to a more comprehensive study in FBM. The external factors such as economic, geographical and social aspects could be incorporated into the models.

8. Acknowledgement

The authors would like to thank the management of Faculty of Information Technology and Communication, Universiti Teknikal Malaysia Melaka and Faculty of Computing and Information Technology Rabigh, King Abdulaziz University Rabigh for providing facilities and financial support.

9. References

Alexandre, M., Adolfo, C.M. and Iong, B. (2008). *On the concept of e-maintenance: Review and current research*. Reliability Engineering and System Safety. Vol. 93, pp. 1165-1187.

Amik, G. and Deshmukh, S.G. (2006). *Applications and Case Studies: Maintenance management-literature review and directions*. Journal of Quality in Maintenance Engineering, Vol 12, No. 3, pp. 205-238.

Bentley, J.P. (1993). *An Introduction to Reliability and Quality Engineering*. New York: John Wiley and Sons Inc.

Burhanuddin, M.A. (2009). *Decision Support Model in Failure-based Computerized Maintenance Management System for Small and Medium Industries*. Thesis of the Doctorate in Philosophy, Faculty of Computer Science and Information System. Universiti Teknologi Malaysia.

Dhillon, B.S. and Liu, Y. (2006). *Human error in maintenance: a review*. Journal of Quality in Maintenance Engineering. Vol 12, No. 1, pp. 21-36.

Elsayed, (1996). Reliability Engineering. Addision Wesley.

Fernandez, O., *Walmsley, R., Petty, D.J. and Labib, A.W. (2003). A decision support maintenance management system development and implementation*. International Journal of Quality and Reliability Management, Vol.20, No.8, pp. 965-979

Geert, W. and Liliane, P. (2004). *Maintenance concept development: A case study*. International Journal Production Economics. Vol. 89, pp. 395-405.

Geert, W. and Liliane, P. (2007). *CIBOCOF: A framework for industrial maintenance concept development*. International Journal Production Economics. Vol 1, pp. 1-8.

Guy, F., Antonios, P. and Steve, P. (1999). *Information systems development, maintenance and enhancement: findings from a UK study*. International Journal of Information Management. Vol. 19, pp. 319-328.

Imad (2007). *The role of maintenance in improving companies' productivity and profitability*. International Journal Production Economics. Vol. 105, pp. 70-78.

Junaidah, H. (2007). *Information Communicatian Technology Adoption among Small and Mediun Industries in Malaysia*. Intemational Jourmal of Business and Information, Vol 12, No. 2, pp. 221-240.

Kittipong, S. (2008). *A Comprehensive Framework for the Adoption of Techno-Relationship Innovations. Empirical Evidence from eCRM in Manufacturing SMIs*. Thesis of the Doctorate in Philosophy. Umea Universiti, Sweden.

Labib A.W. and Yuniarto, M.N. (2009). *Maintenance Strategies in Changeable Manufacturing*. New York: Springer-Verlag Inc, chapter 19, pp.337-351.

Labib, A.W. (1998a). *A logistics approach to managing the millennium information systems problem*. Logistic Information Management, Vol. 11, No. 5, pp. 285-294.

Labib, A.W. (1998b). *World-class maintenance using a computerized maintenance management system*. Journal of Quality in Maintenance Engineering, Vol. 4, No. 1, pp. 66-75.

Labib, A.W. (2004). *A decision analysis model for maintenance policy selection using a CMMS*. Journal of Quality in Maintenance Engineering, Vol. 10, No. 3, pp. 191-202.

Lam, Y. and Lin, Y.Z. (2004). *A shock model for the maintenance problem of a repairable system*. Computers and Operations Research. Vol. 31, pp. 1807-1820.

Lewis, B.T. (1999). *Facility Manager's Operation and Maintenance Handbook*. McGraw-Hill. New York: R.R. Donelley and Sons Company.

Mirka, K. (2008). *An approach for determining the requirements of computerised maintenance management systems*. Computers in Industry. Vol. 59, pp. 32-40.

Nakajima, S. (1988). *Total Productive Maintenance*. Cambridge, MA. Productive Press.

O'Donoghue, C.D. and Prendergast, J.G. (2004). *Implementation and benefits of introducing a computerised maintenance management system into a textile manufacturing company*. Journal of Material Processing Technology. Vol. 153, pp. 226-232.

Palmer, R.D. (1999). *Maintenance Planning and Scheduling Handbook*. McGraw-Hill. New York: R.R. Donelley and Sons Company.

Pascual, R., Meruane, V. and Rey, P.A. (2009). *On the effect of downtime costs and budget constraint on preventive and replacement policies*. Reliability Engineering and System Safety, Vol. 93, pp. 144-151.

Saleh, A.S. and Ndubisi N.O. (2006). *An Evaluation of SMI Development in Malaysia*. International Review of Business Research Papers, Vol. 2, No. 1, pp. 1-14.

Shamsuddin, A, Masjuki, H. and Zahari, T. (2004). State of implementation of TPM in SMIs: a survey study in Malaysia. *Journal of Quality in Maintenance Engineering*. Vol. 10, No. 2, pp 93-106.

Sherif, Y., Osama, A., Imran, F. and Dennis, R. (2008), *A decision support system for concrete bridge deck maintenance*. Advances in Engineering Software. Vol 39, pp. 202-210.

Williams, K.B. and Sawyer, S.C. (2006). *Using Information Technology*. McGraw-Hill. New York.

Zulkifli, T. (2010). A Hybrid Management Model in Decision Support System for Small and Medium Food Processing Industries. Thesis of the Master of Science Degree, Faculty of Information Technology and Communication. Universiti Teknikal Malaysia Melaka.

Zulkifli, T., Anton, S.P., Burhanuddin, M.A. and Habibullah, A. (2008). *Maintenance Decision Support Fuzzy System in Small and Medium Industries using Decision Making Grid*. The International IEEE Conference on Advanced Computer Theory and Engineering (ICACTE'08). ISBN 978-0-7695-3489-3/08, pp. 680-684. Thailand.

12

Arguing over Goals for Negotiation: Adopting an Assumption-Based Argumentation Decision Support System

Maxime Morge
Université Lille 1
France

1. Introduction

Negotiations occur in procurement, commerce, health and government, among organisations (companies and institutions) and individuals. For instance, electronic procurement (respectively electronic commerce) consists of business-to-business (respectively business-to-customer) purchase and provision of resources or services through the Internet. Typically, organisations and individuals invite bids and negotiate costs, volume discounts or special offers. These negotiations can be (at least partially) delegated to software components in order to reach agreements (semi-)automatically (Jennings et al., 2001). For this purpose, software agents must be associated with stakeholders in negotiations.

In negotiations, participation is voluntary and there is no third party imposing a resolution of conflicts. Participants resolve their conflict by verbal means. The aim for all parties is to "make a deal" while bargaining over their interests, typically seeking to maximise their "good" (welfare), and prepared to concede some aspects while insisting on others. Each side tries to figure out what other sides may want most, or may feel is most important. Since real-world negotiations can be resolved by confronting and evaluating the justifications of different positions, argumentation can support such a process. Logical models of argument (Chesñevar et al., 2000) can be used to support rational decision making by agents, to guide and empower negotiation amongst stakeholders and allow them to reach agreements. With the support of argumentation processes, agents decide which agreements can be acceptable to fulfil the requirements of users and the constraints imposed by interlocutors, taking into account their expertises/preferences and the utilities they assign to situations. This is the reason why many works in the area of Artificial Intelligence focus on computational models of argumentation-based negotiation (Rahwan et al., 2003). Logical models of arguments (e.g. Amgoud & Prade (2009); Bench-Capon & Prakken (2006); Kakas & Moraitis (2003)) can be used to encompass the reasoning of agents engaged in negotiations. However, these approaches do not come with a mechanism allowing interacting agents to concede. Since agents can consider multiple goals which may not be fulfilled all together by a set of non-conflicting decisions, e.g. a negotiation agreement, high-ranked goals must be preferred to low-ranked goals on which agents can concede. In this paper we propose an argumentation-based mechanism for decision-making to concede. Adopting the assumption-based approach of argumentation, we propose here an argumentation framework. It is built upon a logic language which holds statements

representing knowledge, goals, and decisions. Preferences are attached to goals. These concrete data structures consist of information providing the backbone of arguments. Due to the abductive nature of practical commonsense reasoning, arguments are built by reasoning backwards. Moreover, arguments are defined as tree-like structures. Our framework is equipped with a computational counterpart (in the form of a formal mapping from it into a set of assumption-based argumentation frameworks). Indeed, we provide the mechanism for solving a decision problem, modeling the intuition that high-ranked goals are preferred to low-ranked goals which can be withdrawn. Thus, we give a clear semantics to the decisions. In this way, our framework suggests some decisions and provides an interactive and intelligible explanation of this choice. Our implementation, called MARGO, is a tool for multi-attribute qualitative decision-making as required, for instance in agent-based negotiation or in service-oriented agents. In a more practical context, our framework is amenable to industrial applications. In particular, MARGO has been used within the the ArguGRID project[1] for service selection and sevice negotiation.
The paper is organised as follows. Section 2 introduces the basic notions of argumentation in the background of our work. Section 3 defines the core of our proposal, i.e. our argumentation-based framework for decision making. Firstly, we define the framework which captures decision problems. Secondly, we define the arguments. Thirdly, we formalize the interactions amongst arguments in order to define our AF (Argumentation Framework). Finally, we provide the computational counterpart of our framework. Section 4 outlines the implementation of our AF and its usage for service-oriented agents. Finally, section 5 discusses some related works and section 6 concludes with some directions for future work.

2. Background

Our argumentation approach is based on Dung's abstract approach to defeasible argumentation (Dung, 1995). Argumentation is abstractly defined as the interaction amongst arguments, reasons supporting claims, which can be disputed by other arguments. In his seminal work, Dung considers arguments as atomic and abstract entities interacting through a binary relation over these interpreted as "the argument x attacks the argument y". More formally, an abstract argumentation framework (AAF for short) is defined as follows.

Definition 1 (AAF). *An* abstract argumentation framework *is a pair* $\mathtt{aaf} = \langle \mathcal{A}, \mathtt{attacks} \rangle$ *where* $\mathcal{A}$ *is a finite set of arguments and* $\mathtt{attacks} \subseteq \mathcal{A} \times \mathcal{A}$ *is a binary relation over* $\mathcal{A}$*. When* $(\mathtt{a}, \mathtt{b}) \in \mathtt{attacks}$ *, we say that* $\mathtt{a}$ *attacks* $\mathtt{b}$*. Similarly, we say that the set* $\mathtt{S}$ *of arguments attacks* $\mathtt{b}$ *when* $\mathtt{a} \in \mathtt{S}$*.*

This framework is abstract since it specifies neither the nature of arguments nor the semantics of the attack relation. However, an argument can be viewed as a reason supporting a claim which can be challenged by other reasons.
According to this framework, Dung introduces various extension-based semantics in order to analyse when a set of arguments can be considered as collectively justified.

Definition 2 (Semantics). *Let* $\mathtt{aaf} = \langle \mathcal{A}, \mathtt{attacks} \rangle$ *be an abstract argumentation framework. For* $\mathtt{S} \subseteq \mathcal{A}$ *a set of arguments, we say that:*

- $\mathtt{S}$ *is* conflict-free *iff* $\forall \mathtt{a}, \mathtt{b} \in \mathtt{S}$ *it is not the case that* $\mathtt{a}$ *attacks* $\mathtt{b}$*;*

[1] http://www.argugrid.eu

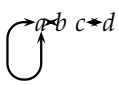
- $\mathtt{S}$ *is* admissible *(denoted* $\mathtt{adm}_{\mathtt{aaf}}(\mathtt{S})$*) iff* $\mathtt{S}$ *is conflict-free and* $\mathtt{S}$ *attacks every argument* $\mathtt{a}$ *such that* $\mathtt{a}$ *attacks some arguments in* $\mathtt{S}$*;*
- $\mathtt{S}$ *is* preferred *iff* $\mathtt{S}$ *is maximally admissible (with respect to set inclusion);*
- $\mathtt{S}$ *is* complete *iff* $\mathtt{S}$ *is admissible and* $\mathtt{S}$ *contains all arguments* $\mathtt{a}$ *such that* $\mathtt{S}$ *attacks all attacks against* $\mathtt{a}$*;*
- $\mathtt{S}$ *is* grounded *iff* $\mathtt{S}$ *is minimally complete (with respect to set inclusion);*
- $\mathtt{S}$ *is* ideal *iff* $\mathtt{S}$ *is admissible and it is contained in every preferred set.*

These declarative model-theoretic *semantics* of the AAF capture various degrees of justification, ranging from very permissive conditions, called *credulous*, to restrictive requirements, called *sceptical*. The semantics of an admissible (or preferred) set of arguments is credulous, in that it sanctions a set of arguments as acceptable if it can successfully dispute every arguments against it, without disputing itself. However, there might be several conflicting admissible sets. That is the reason why various sceptical semantics have been proposed for the AAF, notably the grounded semantics and the sceptically preferred semantics, whereby an argument is accepted if it is a member of all maximally admissible (preferred) sets of arguments. The ideal semantics was not present in (Dung, 1995), but it has been proposed recently (Dung et al., 2007) as a less sceptical alternative than the grounded semantics but it is, in general, more sceptical than the sceptically preferred semantics.

Example 1 (AAF)**.** *In order to illustrate the previous notions, let us consider the abstract argumentation framework* $\mathtt{aaf} = \langle \mathcal{A}, \mathtt{attacks} \rangle$ *where:*

- $\mathcal{A} = \{a, b, c, d\}$*;*
- $\mathtt{attacks} = \{(a,a), (a,b), (b,a), (c,d), (d,c)\}$.

The following graph represents this AAF, whereby the fact that "x $\mathtt{attacks}$ y*" is depicted by a directed arrow from* x *to* y*:*



We can notice that:

- $\{\}$ *is grounded;*
- $\{b, d\}$ *and* $\{b, c\}$ *are preferred.*
- $\{b\}$ *is the maximal ideal set.*

As previously mentionned, Dung's seminal calculus of opposition deals neither with the nature of arguments nor with the semantics of the attacks relation.
Unlike the abstract argumentation, assumption-based argumentation considers neither the arguments nor the attack relations as primitives. Arguments are built by reasoning backwards from conclusions to assumptions given a set of inference rules. Moreover, the attack relation is defined in terms of a notion of "contrary"(Bondarenko et al., 1993; Dung et al., 2007). Actually, assumption-based argumentation frameworks (ABFs, for short) are concrete instances of AAFs built upon deductive systems.
The abstract view of argumentation does not deal with the problem of finding arguments and attacks amongst them. Typically, arguments are built by joining rules, and attacks arise from conflicts amongst such arguments.

Definition 3 (DS)**.** *A deductive system is a pair* $(\mathcal{L}, \mathcal{R})$ *where*

- *$\mathcal{L}$ is a formal language consisting of countably many sentences, and*
- *$\mathcal{R}$ is a countable set of inference rules of the form r: $\alpha \leftarrow \alpha_1, \ldots, \alpha_n$ where $\alpha \in \mathcal{L}$, called the* head *of the rule (denoted $\mathtt{head}(r)$), $\alpha_1, \ldots, \alpha_n \in \mathcal{L}$, called the* body *(denoted $\mathtt{body}(r)$), and $n \geq 0$.*

If $n = 0$, then the inference rule represents an axiom (written simply as α). A deductive system does not distinguish between domain-independent axioms/rules, which belong to the specification of the logic, and domain-dependent axioms/rules, which represents a background theory.
Due to the abductive nature of the practical reasoning, we define and construct arguments by reasoning backwards. Therefore, arguments do not include irrelevant information such as sentences not used to derive a conclusion.

Definition 4 (Deduction). *Given a deductive system $(\mathcal{L}, \mathcal{R})$ and a selection function f, a (backward) deduction of a conclusion α based on a set of premises P is a sequence of sets $S_1, \ldots, S_m$, where $S_1 = \{\alpha\}$, $S_m = \{P\}$, and for every $1 \leq i < m$, where σ is the sentence occurrence in S_i selected by f:*

1. *if σ is not in P then $S_{i+1} = S_i - \{\sigma\} \cup S$ for some inference rule of the form $\sigma \leftarrow S$ in the set of inference rules $\mathcal{R}$;*
2. *if σ is in P then $S_{i+1} = S_i$.*

Deductions are the basis for the construction of arguments in assumption-based argumentation. In order to obtain an argument from a backward deduction, we restrict the premises to those ones that are taken for granted (called *assumptions*). Moreover, we need to specify when one sentence *contraries* an assumptions to specify when one argument attacks another. In this respect, an ABF considers a deductive system augmented by a non-empty set of assumptions and a (total) mapping from assumptions to their contraries. In order to perform decision making, we consider the generalisation of the original assumption-based argumentation framework and the computational mechanisms, whereby multiple contraries are allowed (Gartner & Toni, 2007).

Definition 5 (ABF). *An* assumption-based argumentation framework *is a tuple $\mathtt{abf} = \langle \mathcal{L}, \mathcal{R}, \mathcal{A}sm, \mathcal{C}on \rangle$ where:*

- *$(\mathcal{L}, \mathcal{R})$ is a deductive system;*
- *$\mathcal{A}sm \subseteq \mathcal{L}$ is a non-empty set of* assumptions. *If $x \in \mathcal{A}sm$, then there is no inference rule in $\mathcal{R}$ such that x is the head of this rule;*
- *$\mathcal{C}on$: $\mathcal{A}sm \rightarrow 2^{\mathcal{L}}$ is a (total) mapping from assumptions into set of sentences in $\mathcal{L}$, i.e. their* contraries.

In the remainder of the paper, we restrict ourselves to finite deduction systems, i.e. with finite languages and finite set of rules. For simplicity, we restrict ourselves to flat frameworks (Bondarenko et al., 1993), i.e. whose assumptions do not occur as conclusions of inference rules, such as logic programming or the argumentation framework proposed in this paper.
In the assumption-based approach, the set of assumptions supporting a conclusion encapsulates the essence of the argument.

Definition 6 (Argument). *An* argument *for a conclusion is a deduction of that conclusion whose premises are all assumptions. We denote an argument $\mathtt{a}$ for a conclusion α supported by a set of assumptions $\mathtt{A}$ simply as $\mathtt{a}$: $\mathtt{A} \vdash \alpha$.*

The set of arguments built upon $\mathcal{A}sm$ is denoted $\mathcal{A}(\mathcal{A}sm)$.
In an assumption-based argumentation framework, the attack relation amongst arguments comes from the contrary relation.

Definition 7 (Attack relation). *An argument* $\mathtt{a}$: $\mathtt{A} \vdash \alpha$ attacks *an argument* $\mathtt{b}$: $\mathtt{B} \vdash \beta$ *iff there is an assumption* $x \in \mathtt{B}$ *such as* $\alpha \in \mathcal{C}on(x)$.

According to the two previous definitions, an ABF is clearly a concrete instantiation of an AAF where arguments are deductions and the attack relation comes from the contrary relation.

Example 2 (ABF). *Let* $\mathtt{abf} = \langle \mathcal{L}, \mathcal{R}, \mathcal{A}sm, \mathcal{C}on \rangle$ *be an assumption-based argumentation framework where:*

- $(\mathcal{L}, \mathcal{R})$ *is a deductive system where,*
  - $\mathcal{L} = \{\alpha, \beta, \delta, \gamma, \neg\alpha, \neg\beta, \neg\delta, \neg\gamma\}$,
  - $\mathcal{R}$ *is the following set of rules,*

$$\begin{aligned} \neg\alpha &\leftarrow \alpha \\ \neg\alpha &\leftarrow \beta \\ \neg\beta &\leftarrow \alpha \\ \neg\gamma &\leftarrow \delta \\ \neg\delta &\leftarrow \gamma \end{aligned}$$

- $\mathcal{A}sm = \{\alpha, \beta, \gamma, \delta\}$. *Notice that no assumption is the head of an inference rule in* $\mathcal{R}$;
- *and* $\mathcal{C}on(\alpha) = \{\neg\alpha\}$, $\mathcal{C}on(\beta) = \{\neg\beta\}$, $\mathcal{C}on(\gamma) = \{\neg\gamma\}$, *and* $\mathcal{C}on(\delta) = \{\neg\delta\}$.

Some of the arguments in $\mathtt{abf}$ *are the following:*

$$\begin{aligned} \{\alpha\} &\vdash \neg\alpha \\ \{\alpha\} &\vdash \neg\beta \\ \{\beta\} &\vdash \neg\alpha \\ \{\gamma\} &\vdash \neg\delta \\ \{\delta\} &\vdash \neg\gamma \end{aligned}$$

As stated in Dung et al. (2007), this ABF is a concrete instance of the AAF example proposed previously.

3. Proposal

This section presents our framework to perform decision making. Taking into account its goals and preferences, an agent needs to solve a decision-making problem where the decision amounts to an alternative it can select even if some goals cannot be reached. This agent uses argumentation in order to assess the suitability of alternatives and to identify "optimal" ones. It argues internally to link the alternatives, their features and the benefits that these features guarantee under possibly incomplete knowledge.
We present here the core of our proposal, i.e. an argumentation framework for decision making. Section 3.1 introduces the walk-through example. Section 3.2 introduces the framework used to capture decision problems. Section 3.3 defines the arguments. Section 3.4 defines the interactions amongst our arguments. Section 3.5 defines our AF. Finally, Section 3.6 presents its computational counterpart.

3.1 Walk-through example

We consider e-procurement scenarios where buyers seek to purchase earth observation services from sellers (Stournaras, 2007). Each agent represents a user, i.e. a service requester or a service provider. The negotiation of the fittest image is a complex task due to the number of possible choices, their characteristics and the preferences of the users. Therefore, this usecase is interesting enough for the evaluation of our argumentation-based mechanism for decision-making (Bromuri et al., 2009; Morge & Mancarella, 2010). For simplicity, we abstract away from the real world data of these features and we present here an intuitive and illustrative scenario.

In our scenario, we consider a `buyer` that seeks to purchase a service $\mathtt{s}(x)$ from a `seller`. The latter is responsible for the four following concrete instances of services: `s(a)`, `s(b)`, `s(c)` and `s(d)`. These four concrete services reflect the combinations of their features (cf Tab. 1). For instance, the price of `s(a)` is high (`Price(a,high)`), its resolution is low (`Resolution(a,low)`) and its delivery time is high (`DeliveryTime(a,high)`). According to the preferences and the constraints of the user represented by the `buyer`: the cost must be low (`cheap`); the resolution of the service must be high (`good`); and the delivery time must be low (`fast`). Additionally, the `buyer` is not empowered to concede about the delivery time but it can concede indifferently about the resolution and/or the cost. According to the preferences and constraints of the user represented by the `seller`: the cost of the service must be high; the resolution of the service must be low; and the delivery time must be high (`slow`). The `seller` is not empowered to concede about the cost but it can concede indifferently about the resolution or/and the delivery time. The agents attempt to come to an agreement on the contract for the provision of a service $\mathtt{s}(x)$. Taking into account some goals, preferences and constraints, the `buyer` (resp. the `seller`) needs to interactively solve a decision-making problem where the decision amounts to a service it can buy (resp. provide).

Service	Price	Resolution	DeliveryTime
s(a)	high	low	high
s(b)	high	high	high
s(c)	high	low	low
s(d)	low	low	low

Table 1. The four concrete services and their features

The decision problem of the `buyer` can be captured by an abstract argumentation framework which contains the following arguments:

- d_1 - He will buy `s(d)` if the seller accepts it since the cost is low;
- d_2 - He will buy `s(d)` if the seller accepts it since the delivery time is low;
- c - He will buy `s(c)` if the seller accepts it since the delivery time is low;

Due to the mutual exclusion between the alternatives, c attacks d_1, c attacks d_2, d_1 attacks c and d_2 attacks c. We will illustrate our concrete argumentation framework for decision making with the decision problem of the buyer.

3.2 Decision framework

Since we want to provide a computational model of argumentation for decision making and we want to instantiate it for particular problems, we need to specify a particular language, allowing us to express statements about the various different entities involved

in the knowledge representation for decision making. In our framework, the knowledge is represented by a logical theory built upon an underlying logic-based language.
In this language we distinguish between several different categories of predicate symbols. First of all, we use *goals* to represent the possible objectives of the decision making process. For instance, the goal `fast` represents the objective of a buyer who would like to obtain a quick answer. We will denote by $\mathcal{G}$ the set of predicate symbols denoting goals.
In the language we also want to distinguish symbols representing the *decisions* an agent can adopt. For instance, in the procurement example a unary predicate symbol $\mathtt{s}(x)$ can be used to represent the decision of the buyer to select the service x. It is clear that a problem may involve some decisions over different items, which will correspond to adopting many decision predicate symbols (this is not the case in our running example). We will denote by $\mathcal{D}$ the set of the predicate symbols for representing decisions.
In order to represent further knowledge about the domain under consideration, we will adopt also a set of predicate symbols for *beliefs*, denoted by $\mathcal{B}$. Furthermore, in many situations the knowledge about a decision making problem may be incomplete, and it may require to make assumptions to carry on the reasoning process. This will be tackled by selecting, in the set $\mathcal{B}$, those predicate symbols representing *presumptions* (denoted by $\mathcal{P}sm$). For instance, in the procurement example, the decision made by the buyer may (and will indeed) depend upon the way the buyer thinks the seller replies to the buyer's offer, either by accepting or by rejecting it. This can be represented by a presumption $\mathtt{reply}(x)$, where x is either `accept` or `reject`.
In a decision making problem, we need to express *preferences* between different goals and the *reservation value*, that is the lowest (in terms of preference) set of goals under which the agent cannot concede. For instance, in the procurement example, the buyer prefers to minimize the price. Hence, its knowledge base should somehow represent the fact that the goal `fast` should be preferred to `cheap`. On the other hand, the buyer is prepared to concede on the price in order to achieve an agreement with the seller, but it may be not ready to concede on the delivery time which must be low. Hence, its knowledge base should somehow represent the fact that these goals consist of its reservation value.
Finally, we allow the representation of explicit *incompatibilities* between goals and/or decisions. For instance, different alternatives for the same decision predicate are incompatible with each other, e.g. $\mathtt{s}(\mathtt{a})$ is incompatible with $\mathtt{s}(\mathtt{b})$. On the other hand, different goals *may be* incompatible with one another. For instance, `cheap` is incompatible with `expensive`, whereas `expensive` is not incompatible with `good`. Incompatibilities between goals and between decisions will be represented through a binary relation denoted by $\mathcal{I}$.
The above informal discussion can be summarized by the definition of *decision framework* (Definition 8 below). For the sake of simplicity, in this definition, as well as in the rest of the paper, we will assume some familiarity with the basic notions of logic languages (such as terms, atomic formulae, clauses etc.) Moreover, we will not explicitly introduce formally all the components of the underlying logic language, in order to focus our attention to those components which are relevant to our decision making context. So, for instance, we assume that the constants and function symbols over which terms are built (i.e. predicate arguments) are given. Finally, given a set of predicate symbols X in the language, we will still use X to denote the set of all possible atomic formulae built on predicates belonging to X. If not clear from the context, we will point out whether we refer to the predicate symbols in X rather than to the atomic formulae built on X.

Definition 8 (Decision framework)**.** *A* decision framework *is a tuple* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$*, where:*

- $\mathcal{DL} = \mathcal{G} \cup \mathcal{D} \cup \mathcal{B}$ *is a set of predicate symbols called the* ***decision language****, where we distinguish between goals* ($\mathcal{G}$)*, decisions* ($\mathcal{D}$) *and beliefs* ($\mathcal{B}$)*;*
- $\mathcal{P}sm$ *is a set of atomic formulae built upon predicates in* $\mathcal{DL}$ *called* ***presumptions****;*
- $\mathcal{I}$ *is the* ***incompatibility relation****, i.e. a binary relation over atomic formulae in* $\mathcal{G}$*,* $\mathcal{B}$ *or* $\mathcal{D}$*.* $\mathcal{I}$ *is not necessarily symmetric;*
- $\mathcal{T}$ *is a logic* ***theory*** *built upon* $\mathcal{DL}$*; statements in* $\mathcal{T}$ *are clauses, each of which has a distinguished* name*;*
- $\mathcal{P} \subseteq \mathcal{G} \times \mathcal{G}$ *is the* ***priority*** *relation, namely a transitive, irreflexive and asymmetric relation over atomic formulae in* $\mathcal{G}$*;*
- $\mathcal{RV}$ *is a set of literals built upon predicates in* $\mathcal{G}$*, called the* ***reservation value****.*

Let us summarize the intuitive meaning of the various components of the framework. The language $\mathcal{DL}$ is composed by:

- the set of *goal* predicates, i.e. some predicate symbols which represent the features that a decision must exhibit;
- the set $\mathcal{D}$ of *decision* predicates, i.e. some predicate symbols which represent the actions which must be performed or not; different atoms built on $\mathcal{D}$ represent different *alternatives*;
- the set $\mathcal{B}$ of *beliefs*, i.e. some predicate symbols which represent epistemic statements;.

In this way, we can consider multiple objectives which may or not be fulfilled by a set of decisions under certain circumstances.
We explicitly distinguish *presumable* (respectively *non-presumable*) literals which can (respectively cannot) be assumed to hold, as long as there is no evidence to the contrary. Decisions as well as some beliefs can be assumed. In this way, DF can model the incompleteness of knowledge.
The most natural way to represent conflicts in our object language is by means of some forms of logical negation. We consider two types of negation, as usual, e.g., in extended logic programming, namely *strong negation* $\neg$ (also called *explicit* or *classical negation*), and *weak negation* $\sim$, also called *negation as failure*. As a consequence we will distinguish between strong literals, i.e. atomic formulae possibly preceded by strong negation, and weak literals, i.e. literals of the form $\sim L$, where L is a strong literal. The intuitive meaning of a strong literal $\neg L$ is "L is definitely not the case", while $\sim L$ intuitively means "There is no evidence that L is the case".
The set $\mathcal{I}$ of incompatibilities contains some *default* incompatibilities related to negation on the one hand, and to the nature of decision predicates on the other hand. Indeed, given an atom A, we have $A \, \mathcal{I} \, \neg A$ as well as $\neg A \, \mathcal{I} \, A$. Moreover, $L \, \mathcal{I} \, \sim L$, whatever L is, representing the intuition that L is evidence to the contrary of $\sim L$. Notice, however, that we do not have $\sim L \, \mathcal{I} \, L$, as in the spirit of weak negation. Other default incompatibilities are related to decisions, since different alternatives for the same decision predicate are incompatible with one another. Hence, $D(a_1) \, \mathcal{I} \, D(a_2)$ and $D(a_2) \, \mathcal{I} \, D(a_1)$, D being a decision predicate in $\mathcal{D}$, and a_1 and a_2 being different constants representing different[2] alternatives for D. Depending on the particular decision problem being represented by the framework, $\mathcal{I}$ may contain further

[2] Notice that in general a decision can be addressed by more than two alternatives.

non-default incompatibilities. For instance, we may have $g \mathcal{I} g'$, where g, g' are different goals (as `cheap` $\mathcal{I}$ `expensive` in the procurement example). To summarize, the incompatibility relation captures the conflicts, either default or domain dependent, amongst decisions, beliefs and goals.
The incompatibility relation can be easily lifted to set of sentences. We say that two sets of sentences Φ_1 and Φ_2 are *incompatible* (still denoted by $\Phi_1 \mathcal{I} \Phi_2$) iff there is a sentence ϕ_1 in Φ_1 and a sentence ϕ_2 in Φ_2 such that $\phi_1 \mathcal{I} \phi_2$.
A theory gathers the statements about the decision problem.

Definition 9 (Theory). *A* theory $\mathcal{T}$ *is an extended logic program, i.e a finite set of rules R:* $L_0 \leftarrow L_1, \ldots, L_j, \sim L_{j+1}, \ldots, \sim L_n$ *with* $n \geq 0$, *each* L_i *(with* $i \geq 0$*) being a strong literal in* $\mathcal{L}$. *R, called the unique* name *of the rule, is an atomic formula of* $\mathcal{L}$. *All variables occurring in a rule are implicitly universally quantified over the whole rule. A rule with variables is a scheme standing for all its ground instances.*

Considering a decision problem, we distinguish:

- *goal rules* of the form *R*: $G_0 \leftarrow G_1, \ldots, G_n$ with $n > 0$, where each G_i ($i \geq 0$) is a goal literal in $\mathcal{DL}$ (or its strong negation). According to this rule, the goal G_0 is promoted (or demoted) by the combination of the goal literals in the body;
- *epistemic rules* of the form *R*: $B_0 \leftarrow B_1, \ldots, B_n$ with $n \geq 0$, where each B_i ($i \geq 0$) is a belief literal of $\mathcal{DL}$. According to this rule, B_0 is true if the conditions $B_1, \ldots, B_n$ are satisfied;
- *decision rules* of the form *R*: $G \leftarrow D_1(a_1), \ldots, D_m(a_m), B_1, \ldots, B_n$ with $m \geq 1, n \geq 0$. The head of the rule is a goal (or its strong negation). The body includes a set of decision literals ($D_i(a_i) \in \mathcal{L}$) and a (possibly empty) set of belief literals. According to this rule, the goal is promoted (or demoted) by the decisions $\{D_1(a_1), \ldots, D_m(a_m)\}$, provided that the conditions $B_1, \ldots, B_n$ are satisfied.

For simplicity, we will assume that the names of rules are neither in the bodies nor in the head of the rules thus avoiding self-reference problems. Moreover, we assume that the elements in the body of rules are independent (the literals cannot be deduced from each other), the decisions do not influence the beliefs, and the decisions have no side effects.
Considering statements in the theory is not sufficient to make a decision. In order to evaluate the previous statements, other relevant pieces of information should be taken into account, such as the priority amongst goals. For this purpose, we consider the *priority* relation $\mathcal{P}$ over the goals in $\mathcal{G}$, which is transitive, irreflexive and asymmetric. $G_1 \mathcal{P} G_2$ can be read "G_1 has priority over G_2". There is no priority between G_1 and G_2, either because G_1 and G_2 are *ex æquo* (denoted $G_1 \simeq G_2$), or because G_1 and G_2 are not comparable. The priority corresponds to the relative importance of the goals as far as solving the decision problem is concerned. For instance, we can prefer a fast service rather than a cheap one. This preference can be captured by the priority. The reservation is the minimal set of goals which needs to be reached. The reservation value is the least favourable point at which one will accept a negotiated agreement. It would mean the bottom line that one would be prepared to concede.
In order to illustrate the previous notions, we provide here the decision framework related to the problem described in Section 3.1.

Example 3 (Decision framework). *We consider the procurement example which is described in Section 3.1. The buyer's decision problem is captured by a* decision framework
$DF = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *where:*

- *the decision language $\mathcal{DL}$ distinguishes*
  - *a set of* ***goals*** $\mathcal{G} = \{\texttt{cheap}, \texttt{expensive}, \texttt{fast}, \texttt{slow}, \texttt{bad}, \texttt{good}\}$. *This set of literals identifies various goals as the cost (*`cheap` *or* `expensive`*), the quality of service (*`good` *or* `bad`*) and the availability (*`slow` *or* `fast`*),*
  - *a set of* ***decisions*** $\mathcal{D} = \{\texttt{s}(x) \mid x \in \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}\}\}$. *This set of literals identifies different alternatives,*
  - *a set of* ***beliefs****, i.e. a set of literals identifying various features* $\texttt{Price}(x,y)$, $\texttt{Resolution}(x,y)$ *and* $\texttt{DeliveryTime}(x,y)$ *with* $x \in \{\texttt{a}, \texttt{b}, \texttt{c}, \texttt{d}\}$, $y \in \{\texttt{high}, \texttt{low}\}$ *(which means that y is the level of a certain feature of x) and a set of literals identifying the possible replies of the responders* $\{\texttt{reply}(y) \mid y \in \{\texttt{accept}, \texttt{reject}\}\}$;
- *the set of presumptions $\mathcal{P}sm$ contains the possible replies;*
- *the incompatibility relation $\mathcal{I}$ is trivially defined. In particular,*
  $\texttt{reply}(\texttt{accept})\ \mathcal{I}\ \texttt{reply}(\texttt{reject})$,
  $\texttt{reply}(\texttt{reject})\ \mathcal{I}\ \texttt{reply}(\texttt{accept})$, *and*
  $\texttt{s}(x)\ \mathcal{I}\ \texttt{s}(y)$, *with* $x \neq y$
  $\texttt{good}\ \mathcal{I}\ \texttt{bad}$, $\texttt{bad}\ \mathcal{I}\ \texttt{good}$, $\texttt{expensive}\ \mathcal{I}\ \texttt{cheap}$, $\texttt{cheap}\ \mathcal{I}\ \texttt{expensive}$,
  $\texttt{slow}\ \mathcal{I}\ \texttt{fast}$, $\texttt{fast}\ \mathcal{I}\ \texttt{slow}$;
- *the theory $\mathcal{T}$ (whatever the agent is the buyer or the seller) is the set of rules shown in Table 2;*
- *the preferences of the buyer in our example are such that:*
  $\texttt{fast}\,\mathcal{P}\,\texttt{cheap}$ *and* $\texttt{fast}\,\mathcal{P}\,\texttt{good}$;
- *The reservation value of the buyer is defined as:* $\mathcal{RV} = \{\texttt{fast}\}$. *If the agent is the seller, then the reservation value is defined as :* $\mathcal{RV} = \{\texttt{expensive}\}$.

$r_{11}(x)$: $\texttt{expensive} \leftarrow \texttt{s}(x), \texttt{Price}(x,\texttt{high}), \texttt{reply}(\texttt{accept})$
$r_{12}(x)$: $\texttt{cheap} \leftarrow \texttt{s}(x), \texttt{Price}(x,\texttt{low}), \texttt{reply}(\texttt{accept})$
$r_{21}(x)$: $\texttt{good} \leftarrow \texttt{s}(x), \texttt{Resolution}(x,\texttt{high}), \texttt{reply}(\texttt{accept})$
$r_{22}(x)$: $\texttt{bad} \leftarrow \texttt{s}(x), \texttt{Resolution}(x,\texttt{low}), \texttt{reply}(\texttt{accept})$
$r_{31}(x)$: $\texttt{fast} \leftarrow \texttt{s}(x), \texttt{DeliveryTime}(x,\texttt{low}), \texttt{reply}(\texttt{accept})$
$r_{32}(x)$: $\texttt{slow} \leftarrow \texttt{s}(x), \texttt{DeliveryTime}(x,\texttt{high}), \texttt{reply}(\texttt{accept})$
f_{11}: $\texttt{Price}(\texttt{a},\texttt{high}) \leftarrow$
f_{12}: $\texttt{Resolution}(\texttt{a},\texttt{low}) \leftarrow$
f_{13}: $\texttt{DeliveryTime}(\texttt{a},\texttt{high}) \leftarrow$
f_{21}: $\texttt{Price}(\texttt{b},\texttt{high}) \leftarrow$
f_{22}: $\texttt{Resolution}(\texttt{b},\texttt{high}) \leftarrow$
f_{23}: $\texttt{DeliveryTime}(\texttt{b},\texttt{high}) \leftarrow$
f_{31}: $\texttt{Price}(\texttt{c},\texttt{high}) \leftarrow$
f_{32}: $\texttt{Resolution}(\texttt{c},\texttt{low}) \leftarrow$
f_{33}: $\texttt{DeliveryTime}(\texttt{c},\texttt{low}) \leftarrow$
f_{41}: $\texttt{Price}(\texttt{d},\texttt{low}) \leftarrow$
f_{42}: $\texttt{Resolution}(\texttt{d},\texttt{low}) \leftarrow$
f_{43}: $\texttt{DeliveryTime}(\texttt{d},\texttt{low}) \leftarrow$

Table 2. The rules of the agents

Our formalism allows to capture the incomplete representation of a decision problem with presumable beliefs. Arguments are built upon these incomplete statements.

3.3 Arguments

In order to turn the decision framework presented in the previous section into a concrete argumentation framework, we need first to define the notion of argument. Since we want that our AF not only suggests some decisions but also provides an intelligible explanation of them, we adopt a tree-like structure of arguments. We adopt here the tree-like structure for arguments proposed in (Vreeswijk, 1997) and we extend it with presumptions on the missing information.

Informally, an argument is a deduction for a conclusion from a set of presumptions represented as a tree, with conclusion at the root and presumptions at the leaves. Nodes in this tree are connected by the inference rules, with sentences matching the head of an inference rule connected as parent nodes to sentences matching the body of the inference rule as children nodes. The leaves are either presumptions or the special extra-logical symbol $\top$, standing for an empty set of premises. Formally:

Definition 10 (Structured argument). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework. A* structured argument *built upon* $\mathtt{DF}$ *is composed by a conclusion, some premises, some presumptions, and some sentences. These elements are abbreviated by the corresponding prefixes (e.g.* $\mathtt{conc}$ *stands for conclusion). A structured argument* $\bar{A}$ *can be:*

1. *a* hypothetical argument *built upon an unconditional ground statement. If L is either a decision literal or an presumable belief literal (or its strong/weak negation), then the argument built upon a ground instance of this presumable literal is defined as follows:*

$$\begin{aligned} &\mathtt{conc}(\bar{A}) = L, \\ &\mathtt{premise}(\bar{A}) = \emptyset, \\ &\mathtt{psm}(\bar{A}) = \{L\}, \\ &\mathtt{sent}(\bar{A}) = \{L\}. \end{aligned}$$

or

2. *a* built argument *built upon a rule such that all the literals in the body are the conclusion of arguments.*

2.1) If f is a fact in $\mathcal{T}$ *(i.e.* $\mathtt{body}(f) = \top$[3]*), then the* trivial *argument* $\bar{A}$ *built upon this fact is defined as follows:*

$$\begin{aligned} &\mathtt{conc}(\bar{A}) = \mathtt{head}(f), \\ &\mathtt{premise}(\bar{A}) = \{\top\}, \\ &\mathtt{psm}(\bar{A}) = \emptyset, \\ &\mathtt{sent}(\bar{A}) = \{\mathtt{head}(f)\}. \end{aligned}$$

2.2) If r is a rule in $\mathcal{T}$ *with* $\mathtt{body}(r) = \{L_1, \ldots, L_j, \sim L_{j+1}, \ldots, \sim L_n\}$ *and there is a collection of structured arguments* $\{\bar{A}_1, \ldots, \bar{A}_n\}$ *such that, for each strong literal* $L_i \in \mathtt{body}(r)$, $L_i = \mathtt{conc}(\bar{A}_i)$ *with* $i \leq j$ *and for each weak literal* $\sim L_i \in \mathtt{body}(r)$, $\sim L_i = \mathtt{conc}(\bar{A}_i)$ *with* $i > j$*, we define the* tree *argument* $\bar{A}$ *built upon the rule r and the set* $\{\bar{A}_1, \ldots, \bar{A}_n\}$ *of structured arguments as follows:*

$$\begin{aligned} &\mathtt{conc}(\bar{A}) = \mathtt{head}(r), \\ &\mathtt{premise}(\bar{A}) = \mathtt{body}(r), \\ &\mathtt{psm}(\bar{A}) = \bigcup_{\bar{A}_i \in \{\bar{A}_1, \ldots, \bar{A}_n\}} \mathtt{psm}(\bar{A}_i), \\ &\mathtt{sent}(\bar{A}) = \mathtt{body}(r) \cup \{\mathtt{head}(r)\} \cup \bigcup_{\bar{A}_i \in \{\bar{A}_1, \ldots, \bar{A}_n\}} \mathtt{sent}(\bar{A}_i). \end{aligned}$$

[3] $\top$ denotes the unconditionally true statement.

The set of structured arguments $\{\bar{A}_1, \ldots, \bar{A}_n\}$ *is denoted by* $\texttt{sbarg}(\bar{A})$*, and its elements are called the* subarguments *of* $\bar{A}$*.*

The set of arguments built upon $\texttt{DF}$ *is denoted by* $\mathcal{A}(\texttt{DF})$*.*

Notice that the subarguments of a tree argument concluding the weak literals are hypothetical arguments. Indeed, the conclusion of an hypothetical argument could be a strong or a weak literal, while the conclusion of a built argument is a strong literal. As in (Vreeswijk, 1997), we consider composite arguments, called *tree* arguments, and atomic arguments, called *trivial* arguments. Unlike the other definitions of arguments (set of assumptions, set of rules), our definition considers that the different premises can be challenged and can be supported by subarguments. In this way, arguments are intelligible explanations. Moreover, we consider *hypothetical* arguments which are built upon missing information or a suggestion, i.e. a decision. In this way, our framework allows to reason further by making suppositions related to the unknown beliefs and over possible decisions.
Let us consider the previous example.

Example 4 (Arguments)**.** *The arguments* $\bar{D}_2$ *and* $\bar{C}$ *concluding* $\texttt{fast}$ *are depicted in Fig. 1 and Fig. 2, respectively. They are arguments concluding that the availability is promoted since the delivery time of the services* $\texttt{c}$ *and* $\texttt{d}$ *is low. For this purpose we need to suppose that the seller's reply will be an acceptance. An argument can be represented as a tree where the root is the conclusion (represented by a triangle) directly connected to the premises (represented by losanges) if they exist, and where leaves are either decisions/presumptions (represented by circles) or the unconditionally true statement. Each plain arrow corresponds to a rule (or a fact) where the head node corresponds to the head of the rule and the tall nodes represent the literals in the body of the rule. The tree arguments* $\bar{C}$ *and* $\bar{D}_2$ *are composed of three subarguments: two hypothetical and one trivial argument. Neither trivial arguments nor hypothetical arguments contain subarguments.*

3.4 Interactions between arguments

In order to turn the decision framework into an argumentation framework, we need to capture the interactions between arguments. The interactions amongst structured arguments may come from their conflicts and from the priority over the goals which are promoted by these arguments. We examine in turn these different sources of interaction. Firstly, we define the attack relation amongst conflicting structured arguments in the same way we have defined the attack relation in the assumption-based argumentation frameworks. Secondly, we define the strength of arguments. Finally, we define the defeat relation amongst the structured arguments to capture the whole of interactions amongst them.
Since their sentences are conflicting, the structured arguments interact with one another. For this purpose, we define the following attack relation.

Definition 11 (Attack relation)**.** *Let* $\texttt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework, and* $\bar{A}, \bar{B} \in \mathcal{A}(\texttt{DF})$ *be two structured arguments.* $\bar{A}$ $\texttt{attacks}$ $\bar{B}$ *iff* $\texttt{sent}(\bar{A})\ \mathcal{I}\ \texttt{sent}(\bar{B})$*.*

This relation encompasses both the direct (often called *rebuttal*) attack due to the incompatibility of the conclusions, and the indirect (often called *undermining*) attack, i.e. directed to a "subconclusion". According to this definition, if an argument attacks a subargument, the whole argument is attacked.
Let us go back to our example.

Example 5 (Attack relation)**.** $\bar{D}_2$ *(respectively* $\bar{C}$*) is built upon the hypothetical subargument supposing* $\texttt{s}(\texttt{d})$ *(respectively* $\texttt{s}(\texttt{c})$*). Therefore,* $\bar{C}$ *and* $\bar{D}$ *attack each other.*

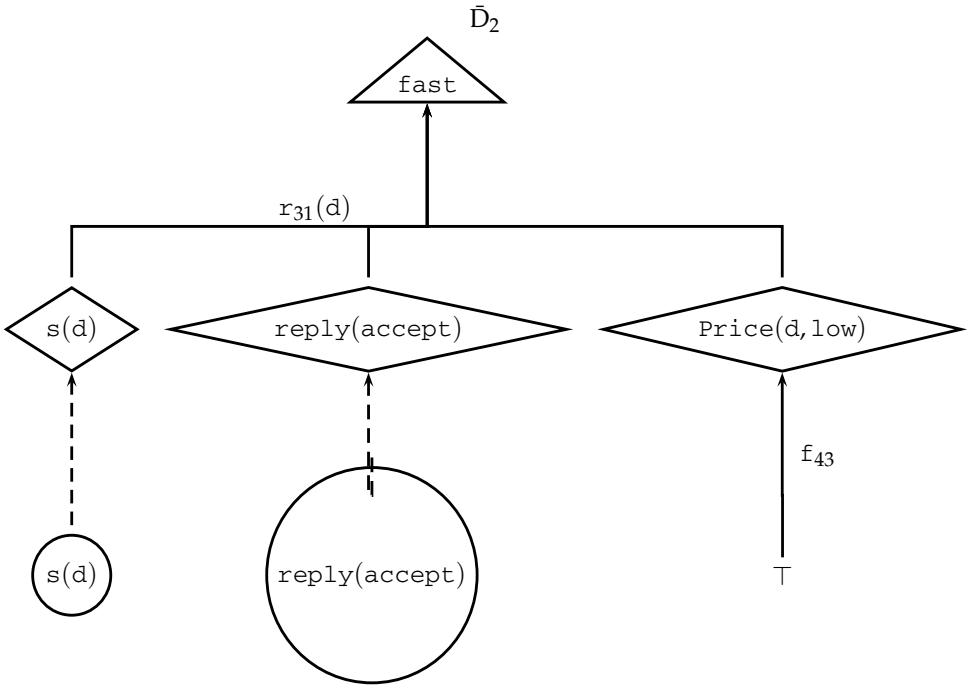


Fig. 1. The argument $\bar{D}_2$ concluding `fast`

Arguments are *concurrent* if their conclusions are identical or incompatible. In order to compare the strength of concurrent arguments, various domain-independent principles of commonsense reasoning can be applied. According to the specificity principle (Simari & Loui, 1992), the most specific argument is stronger one. According to the weakest link principle (Amgoud & Cayrol, 2002), an argument cannot be justified unless all of its subarguments are justified. In accordance with the last link principle (Prakken & Sartor, 1997), the strength of our arguments comes from the preferences between the sentence of the arguments. By contrast, the strength of our argument does not depend on the quality of information used to build that argument but it is determinated by its conclusion.

Definition 12 (Strength relation). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework and* $\bar{A}_1, \bar{A}_2 \in \mathcal{A}(\mathtt{DF})$ *be two structured which are concurrent.* $\bar{A}_1$ *is stronger than* $\bar{A}_2$ *(denoted* $\bar{A}_1 \mathcal{P} \bar{A}_2$*) iff* $\mathtt{conc}(\bar{A}_1) = g_1 \in \mathcal{G}$, $\mathtt{conc}(\bar{A}_2) = g_2 \in \mathcal{G}$ *and* $g_1 \mathcal{P} g_2$.

Due to the definition of $\mathcal{P}$ over $\mathcal{T}$, the relation $\mathcal{P}$ is transitive, irreflexive and asymmetric over $\mathcal{A}(\mathtt{DF})$.
The attack relation and the strength relation can be combined. As in (Amgoud & Cayrol, 1998; Bench-Capon, 2002), we distinguish between one argument attacking another, and that attack succeeding due to the strength of arguments.

Definition 13 (Defeat relation). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework and* $\bar{A}$ *and* $\bar{B}$ *be two structured arguments.* $\bar{A}$ defeats $\bar{B}$ *iff:*

1. $\bar{A}$ `attacks` $\bar{B}$;
2. *and it is not the case that* $\bar{B} \mathcal{P} \bar{A}$.

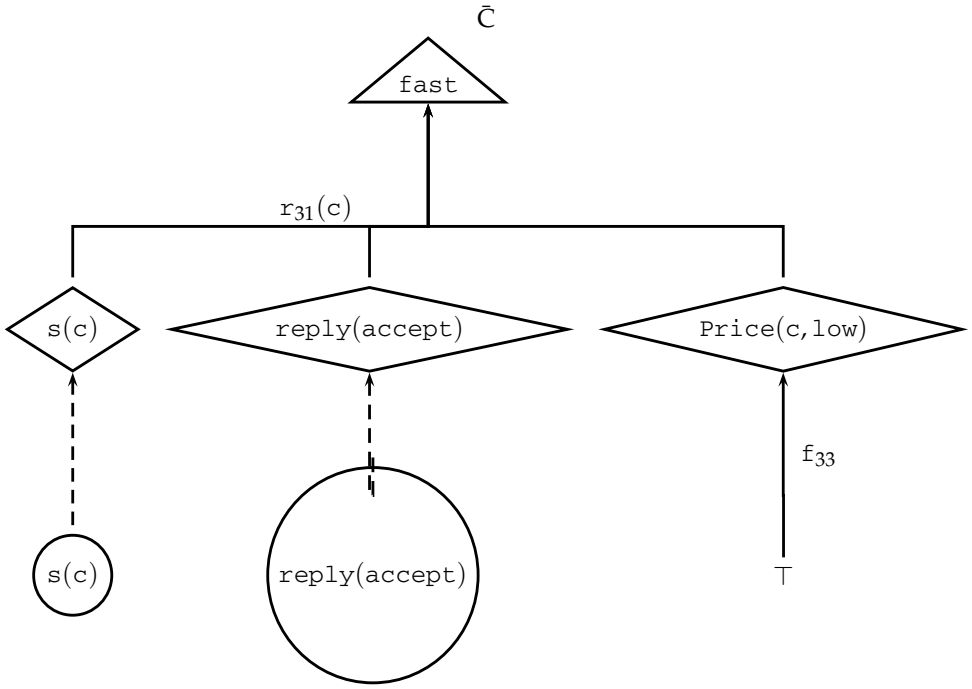


Fig. 2. The argument $\bar{C}$ concluding `fast`

Similarly, we say that a set S of structured arguments defeats a structured argument $\bar{A}$ if $\bar{A}$ is defeated by one argument in S.

Let us consider our example.

Example 6 (Defeat relation)**.** *As previously mentionned, $\bar{C}$ and $\bar{D}_2$ attack each other and they conclude the same goal* `fast`*. We can deduce that $\bar{C}$ and $\bar{D}_2$ defeat each other.*

3.5 Argumentation framework

We are now in the position of summarizing what is our argumentation framework for decision making. In doing this, we also inherit the semantics defined by Dung to analyse when a decision can be considered as acceptable.

As we have seen, in our argumentation-based approach for decision making, arguments motivate decisions and they can be defeated by other arguments. More formally, our argumentation framework (AF for short) is defined as follows.

Definition 14 (AF)**.** *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework. The* argumentation framework for decision making *built upon* `DF` *is a pair* $\mathtt{AF} = \langle \mathcal{A}(\mathtt{DF}), \mathtt{defeats} \rangle$ *where* $\mathcal{A}(\mathtt{DF})$ *is the finite set of structured arguments built upon* `DF` *as defined in Definition 8, and* $\mathtt{defeats} \subseteq \mathcal{A}(\mathtt{DF}) \times \mathcal{A}(\mathtt{DF})$ *is the binary relation over* $\mathcal{A}(\mathtt{DF})$ *as defined in Definition 13.*

We adapt Dung's extension-based semantics in order to analyse whenever a set of structured arguments can be considered as subjectively justified with respect to the preferences.

Definition 15 (Semantics). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework and* $\mathtt{AF} = \langle \mathcal{A}(\mathtt{DF}), \mathtt{defeats} \rangle$ *be our argumentation framework for decision making. For* $\bar{S} \subseteq \mathcal{A}(\mathtt{DF})$ *a set of structured arguments, we say that:*

- $\bar{S}$ *is* subjectively conflict-free *iff* $\forall \bar{A}, \bar{B} \in \bar{S}$ *it is not the case that* $\bar{A}$ *defeats* $\bar{B}$*;*
- $\bar{S}$ *is* subjectively admissible *(s-admissible for short), denoted* $\mathtt{sadm}_{\mathtt{AF}}(\bar{S})$*, iff* $\bar{S}$ *is subjectively conflict-free and* $\bar{S}$ *defeats every argument* $\bar{A}$ *such that* $\bar{A}$ *defeats some argument in* $\bar{S}$*;*

We restrict ourselves to the subjective admissibility, but the other Dung's extension-based semantics (cf Definition 2) can be easily adapted.
Formally, given a structured argument $\bar{\mathrm{A}}$, let

$$\mathtt{dec}(\bar{\mathrm{A}}) = \{D(a) \in \mathtt{psm}(\bar{\mathrm{A}}) \mid D \text{ is a decision predicate}\}$$

be the set of decisions supported by the structured argument $\bar{\mathrm{A}}$.
The decisions are *suggested* to reach a goal if they are supported by a structured argument concluding this goal and this argument is a member of an s-admissible set of arguments.

Definition 16 (Credulous decisions). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework,* $g \in \mathcal{G}$ *be a goal and* $\mathtt{D} \subseteq \mathcal{D}$ *be a set of decisions. The decisions* $\mathtt{D}$ ***credulously argue for*** g *iff there exists an argument* $\bar{A}$ *in a s-admissible set of arguments such that* $\mathtt{conc}(\bar{A}) = g$ *and* $\mathtt{dec}(\bar{A}) = \mathtt{D}$*. We denote* $\mathtt{val}_c(\mathtt{D})$ *the set of goals in* $\mathcal{G}$ *for which the set of decisions* $\mathtt{D}$ *credulously argues.*

It is worth noticing that the decisions which credulously argue for a goal cannot contain mutual exclusive alternatives for the same decision predicate. This is due to the fact that a s-admissible set of arguments is subjectively conflict-free.
If we consider the structured arguments $\bar{\mathrm{A}}$ and $\bar{\mathrm{B}}$ supporting the decisions $D(a)$ and $D(b)$ respectively where a and b are mutually exclusive alternatives, we have $D(a)\ \mathcal{I}\ D(b)$ and $D(a)\ \mathcal{I}\ D(b)$ and so, either $\bar{\mathrm{A}}$ $\mathtt{defeats}$ $\bar{\mathrm{B}}$ or $\bar{\mathrm{B}}$ $\mathtt{defeats}$ $\bar{\mathrm{A}}$ or both of them depending on the strength of these arguments.

Proposition 1 (Mutual exclusive alternatives). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework,* $g \in \mathcal{G}$ *be a goal and* $\mathtt{AF} = \langle \mathcal{A}(\mathtt{DF}), \mathtt{defeats} \rangle$ *be the argumentation framework for decision making built upon* $\mathtt{DF}$*. If* $\bar{S}$ *be a s-admissible set of arguments such that, for some* $\bar{A} \in \bar{S}$*,* $g = \mathtt{conc}(\bar{A})$ *and* $D(a) \in \mathtt{psm}(\bar{A})$*, then* $D(b) \in \mathtt{psm}(\bar{A})$ *iff* $a = b$*.*

However, notice that mutual exclusive decisions can be suggested for the same goal through different s-admissible sets of arguments. This case reflects the credulous nature of our semantics.

Definition 17 (Skeptical decisions). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework,* $g \in \mathcal{G}$ *be a goal and* $\mathtt{D}, \mathtt{D}' \subseteq \mathcal{D}$ *be two sets of decisions. The set* $\mathtt{D}$ *of decisions* ***skeptically argue for*** g *iff for all s-admissible set of arguments* $\bar{S}$ *such that for some arguments* $\bar{A}$ *in* $\bar{S}$ $\mathtt{conc}(\bar{A}) = g$*, then* $\mathtt{dec}(\bar{A}) = \mathtt{D}$*. We denote* $\mathtt{val}_s(\mathtt{D})$ *the set of goals in* $\mathcal{G}$ *for which the set of decisions* $\mathtt{D}$ *skeptically argues. The decisions* $\mathtt{D}$ *is* ***skeptically preferred*** *to the decisions* $\mathtt{D}'$ *iff* $\mathtt{val}_s(\mathtt{D}) \mathcal{P} \mathtt{val}_s(\mathtt{D}')$*.*

Due to the uncertainties, some decisions satisfy goals for sure if they skeptically argue for them, or some decisions can possibly satisfy goals if they credulously argue for them. While the first case is required for convincing a risk-averse agent, the second case is enough to

convince a risk-taking agent. Since some ultimatum choices amongst various justified sets of alternatives are not always possible, we will consider in this paper the most "skeptically preferred" decisions.
The decision making process can be described as the cognitive process in which an agent evaluates the alternatives that are available, according to their features, to determine whether and how they satisfy his needs. The principle for decision making we adopt is that higher-ranked goals should be pursued at the expense of lower-ranked goals, and thus choices enforcing higher-ranked goals should be preferred to those enforcing lower-ranked goals. We are in a situation where there is a ranking of individual objects (the preferences between goals) and we need a ranking that involve subsets of these objects (See Barber et al. (2004) for a survey). For this purpose, we adopt the minmax ordering.

Definition 18 (Preferences). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework. We consider* $\mathtt{G}$, $\mathtt{G}'$ *two sets of goals in* $\mathcal{G}$ *and* $\mathtt{D}$, $\mathtt{D}'$ *two sets of decisions in* $\mathcal{D}$. $\mathtt{G}$ *is* ***preferred*** *to* $\mathtt{G}$ *(denoted* $\mathtt{G}\mathcal{P}\mathtt{G}'$*) iff*

1. $\mathtt{G} \supset \mathtt{G}'$, *and*
2. $\forall g \in \mathtt{G} \setminus \mathtt{G}'$ *there is no* $g' \in \mathtt{G}'$ *such that* $g'\mathcal{P}g$.

$\mathtt{D}$ *is* ***credulously preferred*** *(respectively* ***skeptically preferred****) to* $\mathtt{D}'$ *(denoted* $\mathtt{D}\mathcal{P}_c\mathtt{D}'$ *and* $\mathtt{D}\mathcal{P}_s\mathtt{D}'$*) iff* $\mathtt{val}_c(\mathtt{D})\mathcal{P}\mathtt{val}_c(\mathtt{D}')$ *(respectively* $\mathtt{val}_s(\mathtt{D})\mathcal{P}\mathtt{val}_s(\mathtt{D}')$*).*

Formally, let
$\mathcal{SAD} = \{\mathtt{D} \mid \mathtt{D} \subseteq \mathcal{D}$ such that $\mathcal{RV} \subseteq \mathtt{val}_s(\mathtt{D})$ and
$\forall \mathtt{D}' \subseteq \mathcal{D}$ it is not the case that $\mathcal{RV} \subseteq \mathtt{val}_s(\mathtt{D}')$ and $\mathtt{val}_s(\mathtt{D}')\ \mathcal{P}\ \mathtt{val}_s(\mathtt{D})\}$
be the set of decisions which can be skeptically accepted by the agent. Additionally, let
$\mathcal{SAG} = \{\mathtt{G} \mid \mathtt{G} \subseteq \mathcal{G}$ such that $\mathtt{G} = \mathtt{val}_s(\mathtt{D})$ with $\mathtt{D} \in \mathcal{SAD}\}$
be the goals which can be skeptically reached by the agent.
As an example of the decision making principle, consider the goals g_0, g_1 and g_2 such that $g_2\mathcal{P}g_1$, $g_2\mathcal{P}g_0$ and $\mathcal{RV} = \{g_0\}$. $\{g_2, g_1, g_0\}$ is preferred to both $\{g_2, g_0\}$, $\{g_2, g_1\}$ whereas $\{g_2, g_0\}$, $\{g_2, g_1\}$ are incomparable and so equally preferred. However, $\{g_2, g_1\}$ cannot be reached by the agent since it does not includes the reservation value.
Let us consider now the buyer's decision problem in the procurement example.

Example 7 (Semantics). *The structured argument* $\bar{C}$ *and* $\bar{D}_2$*, which are depicted in Fig. 1 and Fig. 2, conclude* `fast`*. Actually, the sets of decisions* $\{s(c)\}$ *and* $\{s(d)\}$ *credulously argue for* `fast`*. The decisions* $\{s(d)\}$ *skeptically argue for* `cheap` *and a fortiori credulously argue for it. Therefore,* $\{s(d)\}$ *is a skeptically acceptable set of decisions. The reservation value of the buyer only contains* `fast`*. Therefore,* $\{s(d)\}$ *is skeptically preferred to* $\{s(c)\}$ *and* $\{s(d)\}$ *is a skeptical acceptable set of decisions due to the reservation value and the priority over the goals.*

In our example, there is only one suggested set of decisions.
Since agents can consider multiple objectives which may not be fulfilled all together by a set of non-conflicting decisions, they may have to make some concessions, i.e. surrender previous proposals. Concessions are crucial features of agent-based negotiation. Jeffrey S. Rosenschein and Gilad Zlotkin have proposed a monotonic concession protocol for bilateral negotiations in Rosenschein & Zlotkin (1994). In this protocol, each agent starts from the deal that is best for him and either concedes or stands stills in each round. A (monotonic) concession means that an agent proposes a new deal that is better for the other agent. Differently from Rosenschein & Zlotkin (1994), we do not assume that the agent has an interlocutor and if it does, that it does

not know the preferences of its interlocutors. We say that a decision is a minimal concession whenever there is no other preferred decisions.

Definition 19 (Minimal concession). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework. The decision* $\mathtt{dec} \in \mathcal{D}$ *is a* ***concession*** *with respect to* $\mathtt{dec}' \in \mathcal{D}$ *iff there exists a set of decisions* $\mathtt{D}$ *such that* $\mathtt{dec} \in \mathtt{D}$ *and for all* $\mathtt{D}' \subseteq \mathcal{D}$ *with* $\mathtt{dec}' \in \mathtt{D}'$*, it is not the case that* $\mathtt{D}\mathcal{P}\mathtt{D}'$*. The decision* $\mathtt{dec}$ *is a* ***minimal concession*** *wrt* $\mathtt{dec}'$ *iff it is a concession wrt* $\mathtt{dec}'$ *and there is no* $\mathtt{dec}'' \in \mathcal{D}$ *such that*

- $\mathtt{dec}''$ *is a concession wrt* $\mathtt{dec}'$*, and*
- *there is* $\mathtt{D}'' \subseteq \mathcal{D}$ *with* $\mathtt{dec}'' \in \mathtt{D}''$ *with* $\mathtt{D}''\mathcal{P}\mathtt{D}$*.*

The minimal concessions are computed by the computational counterpart of our argumentation framework.

Example 8 (Minimal concession). *According to the buyer,* $\{\mathtt{s}(c)\}$ *is a minimal concession with respect to* $\{\mathtt{s}(d)\}$*.*

3.6 Computational counterpart

Having defined our argumentation framework for decision making, we need to find a computational counterpart for it. For this purpose, we move our AF to an ABF (cf. Section 2) which can be computed by the dialectical proof procedure of (Dung et al., 2006) extended in (Gartner & Toni, 2007). So that, we can compute the suggestions for reaching a goal. Additionally, we provide the mechanism for solving a decision problem, modeling the intuition that high-ranked goals are preferred to low-ranked goals which can be withdrawn.
The idea is to map our argumentation framework built upon a decision framework into a collection of assumption-based argumentation frameworks, that we call *practical assumption-based argumentation frameworks* (PABFs for short). Basically, for each rule r in the theory we consider the assumption $\sim \mathtt{deleted}(r)$ in the set of possible assumptions. By means of this new predicate, we distinguish in a PABF the several distinct arguments that give rise to the same conclusion. Considering a set of goals, we allow each PABF in the collection to include (or not) the rules whose heads are these goals (or their strong negations). Indeed, two practical assumption-based frameworks in this collection may differ in the set of rules that they adopt. In this way, the mechanism consists of a search in the collection of PABFs.

Definition 20 (PABF). *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework and* $\mathtt{G} \in \mathcal{G}$ *a set of goals such that* $\mathtt{G} \supseteq \mathcal{RV}$*. A* practical assumption-based argumentation framework built upon $\mathtt{DF}$ associated with the goals $\mathtt{G}$ *is a tuple* $\mathtt{pabf}_{DF}(G) = \langle \mathcal{L}_{DF}, \mathcal{R}_{DF}, \mathcal{A}sm_{DF}, \mathcal{C}on_{DF} \rangle$ *where:*

(i) $\mathcal{L}_{DF} = \mathcal{DL} \cup \{\mathtt{deleted}\}$[4]*;*

(ii) $\mathcal{R}_{DF}$*, the set of inference rules, is defined as follows:*

- *For each rule* $r \in \mathcal{T}$*, there exists an inference rule* $R \in \mathcal{R}_{DF}$ *such that* $\mathtt{head}(R) = \mathtt{head}(r)$ *and* $\mathtt{body}(R) = \mathtt{body}(r) \cup \{\sim \mathtt{deleted}(r)\}$*;*
- *If* $r_1, r_2 \in \mathcal{T}$ *with* $\mathtt{head}(r_1)\ \mathcal{I}\ \mathtt{head}(r_2)$ *and it is not the case that* $\mathtt{head}(r_2)\mathcal{P}\mathtt{head}(r_1)$*, then the inference rule*
  $\mathtt{deleted}(r_2) \leftarrow \sim \mathtt{deleted}(r_1)$ *is in* $\mathcal{R}_{DF}$*.*

(iii)$\mathcal{A}sm_{DF}$*, the set of assumptions, is defined such that* $\mathcal{A}sm_{DF} = \Delta \cup \Phi \cup \Psi \cup \Upsilon \cup \Sigma$ *where:*

- $\Delta = \{D(a) \in \mathcal{L} \mid D(a)$ *is a decision literal* $\}$*,*

[4] We assume $\mathtt{deleted} \notin \mathcal{L}$.

- $\Phi = \{B \in \mathcal{B} \mid B \in \mathcal{P}sm\}$,
- $\Psi = \{\sim \texttt{deleted}(r) \mid r \in \mathcal{T}$ *and* $\texttt{head}(r) \in \{L, \neg L\}$ *s.t.* $L \notin \mathcal{G}\}$,
- $\Upsilon = \{\sim \texttt{deleted}(r) \mid r \in \mathcal{T}$ *and* $\texttt{head}(r) \in \{g, \neg g\}$ *s.t.* $g \in \mathcal{RV}\}$;
- $\Sigma = \{\sim \texttt{deleted}(r) \mid r \in \mathcal{T}$ *and* $\texttt{head}(r) \in \{g, \neg g\}$ *s.t.* $g \in \mathcal{G} - \mathcal{RV}\}$;

(iv)$\mathcal{C}on_{\mathtt{DF}}$ *the set of contraries is defined such that for all* $\alpha \in \mathcal{A}sm_{\mathtt{DF}}$, $y \in \mathcal{C}on(\alpha)$ *iff* $y \mathcal{I} \alpha$.

The set of practical assumption-based argumentation frameworks built upon DF *and associated with the goals* $\mathcal{G}'$ *with* $\mathcal{RV} \subseteq \mathcal{G}' \subseteq \mathcal{G}$ *will be denoted* $\mathtt{PABFS}_{\mathtt{DF}}(\mathcal{G})$.

Case *(i)* defines the language. In order to capture the decision problem within an assumption-based argumentation framework, we have extended the decision language to include a predicate symbol deleted, which is used to specify whether or not a rule is adopted within the PABF. It is worth noticing that the definition of arguments in the ABF (cf Definition 6) focuses attention on the candidate assumptions and ignores the internal structure of arguments. In order to distinguish in a PABF the several distinct arguments that give rise to the same conclusion, we have named the rules used to deduce it. Therefore, an argument in a PABF contains a set of assumptions of the following schemata $\sim \texttt{deleted}(r)$, for all rule r used by the argument.
Case *(ii)* defines the inference rules. Firstly, there is an inference rule for each rule of the theory. For this purpose, the body of each rule r is extended by adding the assumption $\sim \texttt{deleted}(r)$. Referring to Example 3, the rule $\texttt{r}_{11}(x)$ becomes
$\texttt{expensive} \leftarrow \texttt{s}(x), \texttt{Price}(x, \texttt{high}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{11}(x))$.
In this way, the assumption $\sim \texttt{deleted}(\texttt{r}_{11}(x))$ allows an argument to use this rule.
Secondly, the inference rules include not only the original deduction rules but also the conflicts amongst the rules having incompatible heads. It is worth noticing that the attack relation between arguments in the ABF (cf Def. 7) ignores the possible conflicts amongst the heads of rules which are not assumptions. In order to capture these conflicts, we have introduced rules which allow the defeasibility of rules. Referring to the example, we introduce, e.g.,
$\texttt{deleted}(\texttt{r}_{12}(x)) \leftarrow \sim \texttt{deleted}(\texttt{r}_{11}(x))$
modeling the given incompatibility $\texttt{cheap}\ \mathcal{I}\ \texttt{expensive}$. Obviously, we also introduce,
$\texttt{deleted}(\texttt{r}_{11}(x)) \leftarrow \sim \texttt{deleted}(\texttt{r}_{12}(x))$
modeling the given incompatibility $\texttt{expensive}\ \mathcal{I}\ \texttt{cheap}$. Our treatment of conflicting rules requires not to interfere with our treatment of priorities which is inspired by (Kowalski & Toni, 1996). Referring to the example, we introduce, e.g.,
$\texttt{deleted}(\texttt{r}_{12}(x)) \leftarrow \sim \texttt{deleted}(\texttt{r}_{31}(x))$
modeling the given priority $\texttt{cheap}\mathcal{P}\texttt{fast}$. In this way, the corresponding literal
$\sim \texttt{deleted}(\texttt{r}_{31}(x))$ must be assumed in order to handle this priority. Obviously, we do not introduce,
$\texttt{deleted}(\texttt{r}_{31}(x)) \leftarrow \sim \texttt{deleted}(\texttt{r}_{12}(x))$.
Case *(iii)* defines the assumptions. The decisions are obviously possible assumptions. In the same way, a PABF adopts an presumable belief if this is a presumption of the corresponding AF. Referring to the example, an argument, which assumes that the reply is an acceptance, can be built since $\texttt{reply}(\texttt{accept}) \in \mathcal{A}sm_{\mathtt{DF}}$. Each framework adopt the epistemic rules, i.e r with $\texttt{head}(r) \in \{L, \neg L\}$ and $L \notin \mathcal{G}$, by having the assumption $\sim \texttt{deleted}(r)$ in its set of assumptions.
We want to go through the set of goals such that high-ranked goals are preferred to low-ranked goals and the reservation value is the minimal set of goals we want to reach. For this purpose, we adopt the rules concluding the goals (or their negation) in the reservation value,

i.e r with $\texttt{head}(r) \in \{g, \neg g\}$ and $g \in \mathcal{RV}$, by having the assumption $\sim \texttt{deleted}(r)$ in its set of assumptions. However, each framework in $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$ can or cannot adopt the rules concluding the goals (or their negation) which are not in the reservation value, i.e. r with $\texttt{head}(r) \in \{g, \neg g\}$ and $g \in \texttt{G} - \mathcal{RV}$ by having the assumption $\sim \texttt{deleted}(r)$ in its set of assumptions. Referring to the running example and considering the goal `cheap` the strongest structured arguments concluding `cheap`, requires $\texttt{r}_{12}(x)$ to be built within the PABF if $\sim \texttt{deleted}(\texttt{r}_{12}(x)) \in \mathcal{A}sm_{\texttt{DF}}$.
Case *(iv)* defines the contrary relation of a PABF which trivially comes from the incompatibility relation and which comes from the contradiction of $\texttt{deleted}(r)$ with $\sim \texttt{deleted}(r)$ whatever the rule r is.
Arguments will be built upon rules, the candidate decisions, and by making suppositions within the presumable beliefs. Formally, given a decision framework `DF` and a practical assumption-based framework
$\texttt{pabf}_{\texttt{DF}}(\texttt{G}) = \langle \mathcal{L}_{\texttt{DF}}, \mathcal{R}_{\texttt{DF}}, \mathcal{A}sm_{\texttt{DF}}, \mathcal{C}on_{\texttt{DF}} \rangle$, we define

$$\Sigma = \{\sim \texttt{deleted}(r) \in \mathcal{A}sm_{\texttt{DF}} \mid r \in \mathcal{T} \text{ and } \texttt{head}(r) \in \{g, \neg g\} \text{ and } g \in \texttt{G} - \mathcal{RV}\}$$

as the set of goal rules considered in this PABF.
The practical assumption-based argumentation frameworks built upon a decision framework and associated with some goals include (or not) the rules concluding these goals which are more or less prior. This allows us to associate the set $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$ with a priority relation, denoted $\mathcal{P}$, modeling the intuition that, in solving a decision problem, high-ranked goals are preferred to low-ranked goals.

Definition 21 (Priority over PABF). *Let $\texttt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ be a decision framework, $\texttt{G} \in \mathcal{G}$ a set of goals such that $\texttt{G} \supseteq \mathcal{RV}$ and $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$ be the set of PABFs associated with the goals $\texttt{G}$.*
$\forall \texttt{G}_1, \texttt{G}_2$ such that $\mathcal{RV} \subseteq \texttt{G}_1, \texttt{G}_2 \subseteq \texttt{G}$ $\forall \texttt{pabf}_{\texttt{DF}}(\texttt{G}_1), \texttt{pabf}_{\texttt{DF}}(\texttt{G}_2) \in \texttt{PABFS}_{\texttt{DF}}(\texttt{G})$, $\texttt{pabf}_{\texttt{DF}}(\texttt{G}_1) \mathcal{P} \texttt{pabf}_{\texttt{DF}}(\texttt{G}_2)$ iff:

- *$\texttt{G}_1 \supset \texttt{G}_2$, and*
- *$\forall g_1 \in \texttt{G}_1 \setminus \texttt{G}_2$ there is no $g_2 \in \texttt{G}_2$ such that $g_2 \mathcal{P} g_1$.*

Due to the properties of set inclusion, the priority relation $\mathcal{P}$ is transitive, irreflexive and asymmetric over $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$.
In order to illustrate the previous notions, let us go back to our example.

Example 9 (PABF). *Given the decision framework (cf example 3) capturing the decision problem of the buyer ($\mathcal{RV} = \{\texttt{fast}\}$). We consider the set of goals $\{\texttt{fast}, \texttt{cheap}, \texttt{good}\}$. We denote this set $\texttt{G}$. We will consider the collection of practical assumption-based argumentation frameworks $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$.*
Let $\texttt{pabf}_{\texttt{DF}}(\texttt{G}) = \langle \mathcal{L}_{\texttt{DF}}, \mathcal{R}_{\texttt{DF}}, \mathcal{A}sm_{\texttt{DF}}, \mathcal{C}on_{\texttt{DF}} \rangle$ be a practical assumption-based argumentation framework in $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$. This PABF is defined as follows:

- *$\mathcal{L}_{\texttt{DF}} = \mathcal{DL} \cup \{\texttt{deleted}\}$, where $\mathcal{DL}$ is defined as in the previous example and `deleted` specifies if a rule does not hold;*
- *$R_{\texttt{DF}}$ is defined by the rules in Table 3;*
- *$\mathcal{A}sm_{\texttt{DF}} = \Delta \cup \Gamma \cup \Upsilon \cup \Sigma$ where:*
  - $\Delta = \{s(x) \mid x \in \{a, b, c, d\}\}$,
  - $\Phi = \{\texttt{reply}(y) \mid y \in \{\texttt{accept}, \texttt{reject}\}\}$,

- $\Psi = \{ \sim \texttt{deleted}(f_{11}), \sim \texttt{deleted}(f_{12}), \sim \texttt{deleted}(f_{13}) \sim \texttt{deleted}(f_{21}), \sim \texttt{deleted}(f_{22}), \sim \texttt{deleted}(f_{23}), \sim \texttt{deleted}(f_{31}), \sim \texttt{deleted}(f_{32}), \sim \texttt{deleted}(f_{33}), \sim \texttt{deleted}(f_{41}), \sim \texttt{deleted}(f_{42}), \sim \texttt{deleted}(f_{43}) \}$,
- $\Upsilon = \{\sim \texttt{deleted}(r_{31})(x), \sim \texttt{deleted}(r_{32})(x)\}$,
- $\Sigma \subseteq \{\sim \texttt{deleted}(r_{11}(x)), \sim \texttt{deleted}(r_{12}(x)), \sim \texttt{deleted}(r_{21}(x)), \sim \texttt{deleted}(r_{22}(x))\}$;

- *$\mathcal{C}on_{DF}$ is defined trivially. In particular,*
  $\mathcal{C}on(s(x)) = \{s(y) \mid y \neq x\}$,
  for each r, $\texttt{deleted}(r) \in \mathcal{C}on(\sim \texttt{deleted}(r))$ *if* $\sim \texttt{deleted}(r) \in \mathcal{A}sm_{DF}$.

The possible sets Σ considered for the definition of the practical assumption-based argumentation framework $\texttt{pabf}_{DF}(G_i) \in \texttt{PABFS}_{DF}(G)$ (with $1 \leq i \leq 6$) are such that:

- $G_1 = \{\texttt{cheap}, \texttt{good}, \texttt{fast}\}$ *with*
  $\Sigma_1 = \{\sim \texttt{deleted}(r_{11}(x)), \sim \texttt{deleted}(r_{12}(x)), \sim \texttt{deleted}(r_{21}(x)), \sim \texttt{deleted}(r_{22}(x)), \sim \texttt{deleted}(r_{31}(x)), \sim \texttt{deleted}(r_{32}(x))\}$;
- $G_2 = \{\texttt{good}, \texttt{fast}\}$ *with*
  $\Sigma_2 = \{\sim \texttt{deleted}(r_{21}(x)), \sim \texttt{deleted}(r_{22}(x)), \sim \texttt{deleted}(r_{31}(x)), \sim \texttt{deleted}(r_{32}(x))\}$;
- $G_3 = \{\texttt{good}, \texttt{fast}\}$ *with*
  $\Sigma_3 = \{\sim \texttt{deleted}(r_{21}(x)), \sim \texttt{deleted}(r_{22}(x)), \sim \texttt{deleted}(r_{31}(x)), \sim \texttt{deleted}(r_{32}(x))\}$;
- $G_4 = \{\texttt{fast}\}$ *with*
  $\Sigma_4 = \{\sim \texttt{deleted}(r_{31}(x)), \sim \texttt{deleted}(r_{32}(x))\}$.

It is clear that $\texttt{pabf}_{\texttt{DF}}(\texttt{G}_1)\mathcal{P}\texttt{pabf}_{\texttt{DF}}(\texttt{G}_2)$, $\texttt{pabf}_{\texttt{DF}}(\texttt{G}_1)\mathcal{P}\texttt{pabf}_{\texttt{DF}}(\texttt{G}_3)$, $\texttt{pabf}_{\texttt{DF}}(\texttt{G}_2)\mathcal{P}\texttt{pabf}_{\texttt{DF}}(\texttt{G}_4)$ and $\texttt{pabf}_{\texttt{DF}}(\texttt{G}_3)\mathcal{P}\texttt{pabf}_{\texttt{DF}}(\texttt{G}_4)$.
Having defined the PABFs, we show how a structured argument as in Definition 10 corresponds to an argument in one of the PABFs. To do this, we first define a mapping between a structured argument and a set of assumptions.

Definition 22 (Mapping between arguments). *Let $\texttt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ be a decision framework. Let $\bar{A}$ be a structured argument in $\mathcal{A}(\texttt{DF})$ and concluding $\alpha \in \mathcal{DL}$. The corresponding set of assumptions deducing α (denoted $\ell(\bar{A})$) is defined according to the nature of $\bar{A}$.*

- *If $\bar{A}$ is a hypothetical argument, then $\ell(\bar{A}) = \{\alpha\}$.*
- *If $\bar{A}$ is a trivial argument built upon the fact f, then $\ell(\bar{A}) = \{\sim \texttt{deleted}(f)\}$.*
- *If $\bar{A}$ is a tree argument, then*
  $\ell(\bar{A}) = \{\sim \texttt{deleted}(r_1), \ldots, \sim \texttt{deleted}(r_n)\} \cup \{L_1, \ldots, L_m\}$ *where:*
  (i) $r_1, \ldots, r_n$ are the rules of $\bar{A}$;
  (ii) the literals $L_1, \ldots, L_m$ are the presumptions and the decision literals of $\bar{A}$.

The mapping is materialized through a bijection ℓ: $\mathcal{A}(\texttt{DF}) \rightarrow \mathcal{A}sm_{\texttt{DF}}$ where $\mathcal{A}sm_{\texttt{DF}}$ is the set of possible assumptions of one of the PABFs built upon $\texttt{DF}$ and $\mathcal{A}(\texttt{DF})$ is the set of structured arguments built upon $\texttt{DF}$. If $\bar{\texttt{S}}$ is a set of arguments $\mathcal{A}(\texttt{DF})$, we denote $\ell(\bar{\texttt{S}})$ the corresponding set of assumptions. Formally,

$$\ell(\bar{\texttt{S}}) = \{\ell(\bar{\texttt{A}}) \mid \bar{\texttt{A}} \in \bar{\texttt{S}}\}$$

$\texttt{expensive} \leftarrow \texttt{s}(x), \texttt{Price}(x, \texttt{high}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{11}(x))$
$\texttt{cheap} \leftarrow \texttt{s}(x), \texttt{Price}(x, \texttt{low}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{12}(x))$
$\texttt{good} \leftarrow \texttt{s}(x), \texttt{Resolution}(x, \texttt{high}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{21}(x))$
$\texttt{bad} \leftarrow \texttt{s}(x), \texttt{Resolution}(x, \texttt{low}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{22}(x))$
$\texttt{fast} \leftarrow \texttt{s}(x), \texttt{DeliveryTime}(x, \texttt{low}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{31}(x))$
$\texttt{slow} \leftarrow \texttt{s}(x), \texttt{DeliveryTime}(x, \texttt{high}), \texttt{reply}(\texttt{accept}), \sim \texttt{deleted}(r_{32}(x))$
$\texttt{Price}(\texttt{a}, \texttt{high}) \leftarrow \sim \texttt{deleted}(f_{11})$
$\texttt{Resolution}(\texttt{a}, \texttt{low}) \leftarrow \sim \texttt{deleted}(f_{12})$
$\texttt{DeliveryTime}(\texttt{a}, \texttt{high}) \leftarrow \sim \texttt{deleted}(f_{13})$
$\texttt{Price}(\texttt{b}, \texttt{high}) \leftarrow \sim \texttt{deleted}(f_{21})$
$\texttt{Resolution}(\texttt{b}, \texttt{high}) \leftarrow \sim \texttt{deleted}(f_{22})$
$\texttt{DeliveryTime}(\texttt{b}, \texttt{high}) \leftarrow \sim \texttt{deleted}(f_{23})$
$\texttt{Price}(\texttt{c}, \texttt{high}) \leftarrow \sim \texttt{deleted}(f_{31})$
$\texttt{Resolution}(\texttt{c}, \texttt{low}) \leftarrow \sim \texttt{deleted}(f_{32})$
$\texttt{DeliveryTime}(\texttt{c}, \texttt{low}) \leftarrow \sim \texttt{deleted}(f_{33})$
$\texttt{Price}(\texttt{d}, \texttt{low}) \leftarrow \sim \texttt{deleted}(f_{41})$
$\texttt{Resolution}(\texttt{d}, \texttt{low}) \leftarrow \sim \texttt{deleted}(f_{42})$
$\texttt{DeliveryTime}(\texttt{d}, \texttt{low}) \leftarrow \sim \texttt{deleted}(f_{43})$

Table 3. The rules of the PABF

There is a one-to-one mapping between arguments in our AF and arguments in some corresponding PABFs.

Lemma 1 (Mapping between arguments). *Let $\texttt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ be a decision framework, $\texttt{G} \in \mathcal{G}$ a set of goals and $\texttt{PABFS}_{\texttt{DF}}(\texttt{G})$ be the set of PABFs associated with the goals $\texttt{G}$.*

1. *Given a structured argument built upon $\texttt{DF}$ concluding $\alpha \in \mathcal{DL}$, there is a corresponding argument deducing α in some PABFs of $\texttt{PABFS}(\texttt{G})$.*
2. *Given an atomic formula $\alpha \in \mathcal{DL}$ and an argument of a PABF in $\texttt{PABFS}(\texttt{G})$ deducing α, there exists a corresponding structured argument in $\mathcal{A}(\texttt{DF})$ concluding α.*

Let us consider the previous example.

Example 10 (Assumptions). *The arguments in some PABFs corresponding to the structured arguments $\bar{D}_2$ and $\bar{C}$ include the following set of assumptions:*

- $\ell(\bar{D}_2) = \{\sim \texttt{deleted}(r_{31}(\texttt{d})), \sim \texttt{deleted}(f_{43}), \texttt{s}(\texttt{d}), \texttt{reply}(\texttt{accept})\}$;
- $\ell(\bar{C}) = \{\sim \texttt{deleted}(r_{31}(\texttt{c})), \sim \texttt{deleted}(f_{33}), \texttt{s}(\texttt{c}), \texttt{reply}(\texttt{accept})\}$;

Both of them are tree argument. The corresponding set of assumptions $\ell(\bar{D}_2)$ considers the literals $\sim \texttt{deleted}(r_{31}(\texttt{d}))$ and $\sim \texttt{deleted}(f_{43})$ since $\bar{D}_2$ is built upon these rules. Moreover, the literal $\texttt{s}(\texttt{d})$ (respectively $\texttt{reply}(\texttt{accept})$) is a decision literal (respectively a presumption).

In order to compute our extension-based semantics, we explore the collection of PABFs associated to our AF in order to find the PABF which deduces the strongest goals as possible. Indeed, we have developed a mechanism to explore the collection of PABFs associated to our AF in order to compute it. If a s-admissible set of structured arguments concludes some goals, then there is a corresponding admissible set of assumptions in one of the corresponding PABFs and there is no other PABF, where an admissible set of assumptions deduces stronger goals.

Theorem 1 (Mapping between semantics). *Let* $DF = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework and* $AF = \langle \mathcal{A}(DF), \ defeats \rangle$ *be our argumentation framework for decision making. Let us consider* $G \in \mathcal{G}$ *with* $G \supseteq \mathcal{RV}$.

- *If there is a s-admissible set of structured arguments* $\bar{S}_1$ *concluding* G_1 *with* $\mathcal{RV} \subseteq G_1 \subseteq G$ *such there is no s-admissible set of structured arguments concluding* G_2 *with* $\mathcal{RV} \subseteq G_2 \subseteq G$ *with* $G_2 \mathcal{P} G_1$*, then there is* $pabf_1 \in PABFS_{DF}(G)$ *such that the corresponding set of assumptions* $\ell(\bar{S}_1)$ *is admissible within* $pabf_1$ *and there is no* $pabf_2 \in PABFS_{DF}(G)$*, with* $pabf_2 \mathcal{P} pabf_1$*, which contains an admissible set of assumptions deducing* G_2.
- *If there is* $pabf_1 \in PABFS_{DF}(G)$ *which contains an admissible set of assumptions* A_1 *deducing* G_1 *with* $\mathcal{RV} \subseteq G_1 \subseteq G$ *such that there is no* $pabf_2 \in PABFS_{DF}(G)$*, with* $pabf_2 \mathcal{P} pabf_1$*, which contains an admissible set of assumptions deducing* G_2 *with* $\mathcal{RV} \subseteq G_2 \subseteq G$ *and* $G_2 \mathcal{P} G_1$*, then the corresponding structured arguments* $\ell^{-1}(A_1)$ *concluding* G_1 *is in a s-admissible set and there is no other structured arguments* $\bar{S}_2$ *concluding* G_2 *which is in a s-admissible set.*

4. Implementation

The implementation of our framework is called MARGO. We describe here its usage in particular in the context of service-oriented agents. MARGO stands for Multiattribute ARGumentation framework for Opinion explanation. MARGO is written in Prolog and available in GPL (GNU General Public License) at `http://margo.sourceforge.net/`.
In order to be computed by MARGO, the file, which describes the decision problem, contains:

- a set of decisions, i.e. some lists which contain the alternatives courses of actions;
- possibly a set of incompatibilities, i.e. some couples such that the first component is incompatible with the second component;
- possibly a set of symmetric incompatibilities, i.e. some couples such that the first component is incompatible with the second component and conversely;
- a set of decisions rules, i.e. some triples of name - head - body which are simple Prolog representations of the decision rules in our AF;
- possibly a set of goal rules, i.e. some triples of name - head - body which are simple Prolog representations of the goal rules in our AF;
- possibly a set of epistemic rules, i.e. some triples of name - head - body which are simple Prolog representations of the epistemic rules in our AF;
- possibly a set of priorities, some couples of goals such that the former have priority over the latter;
- a set of presumable belief literals;
- a reservation value, i.e. a list which contains the minimal set of goals which needs to be reached.

We can note that the incompatibilities between the mutual exclusive alternatives are implicit in the MARGO language. It is worth noticing that MARGO attempts to narrow the gap between the specification of the decision framework and the corresponding code.
The main predicate `admissible(+G,?AG, ?AD)` succeeds when `AG` are the acceptable goals extracted form `G` and `AD` are the acceptable decisions. The predicate for argument manipulation `admissibleArgument(+C, ?P, ?S)` succeeds when `P` are the premises and

`S` are the presumptions of an argument deriving the conclusion `C` and this argument is in a subjectively admissible set.

Example 11 (Usage). *Figure 3 depicts our example, as described in Section 3.2, in the MARGO syntax.*
`admissible([cheap, fast, good], AG, AD)` *returns:*

```
AG = [cheap,  fast]
AD = [s(d)]
```

`admissibleArgument(cheap,P,S)` *returns:*

```
P = [s(c), price(c,low), reply(accept)]
S = [s(c),reply(accept)]
```

MARGO has been used for service composition and orchestration within the ARGUGRID project[5]. As discussed in (Toni et al., 2008), the ArguGRID system contains a semantic composition environment, allowing users to interact with their agents, and a grid middleware for the actual deployment of services. Service-oriented computing is an interesting test bed for multi-agent system techniques, where agents need to adopt a variety of roles that will empower them to provide services in open and distributed systems. Moreover, service-oriented computing can benefit from multi-agent systems technologies by adopting the coordination mechanisms, interaction protocols, and decision-making tools designed for multi-agent systems, e.g. MARGO.
Bromuri et al. (Bromuri et al., 2009) have demonstrated the use of a fully decentralised multi-agent system supporting agent-automated service discovery, agent-automated service selection, and agent-automated negotiation of Service Level Agreements (SLAs) for the selected services.
Requester agents select services according to their suitability to fulfil high-level user requirements. These agents use argumentation in order to assess suitability and identify "optimal" services. They argue internally using our concrete argumentation system linking decisions on selecting services, (a possibly incomplete description of) the features of these services, the benefits that these features guarantee (under possibly incomplete knowledge). The ArguGRID system uses the MARGO tool for multi-attribute qualitative decision-making to support the decision on suitable services.
As soon as the requester agents identify a suitable service, it engages in a negotiation process with the provider agent for that service. The negotiation aims at agreeing a SLA on the usage of the identified service, and is conducted using a realisation of the minimal concession strategy of (Morge & Mancarella, 2010). According to this, agents start the negotiation with their best offers. During the negotiation, an agent may concede or stand still. It concedes minimally if the other agent has conceded in the previous step or it is making a move in the third step of the negotiation (after offers by both agents have been put forward). It stands still if the other agent has stood still in the previous step. This strategy has useful properties: it is guaranteed to terminate and it is in symmetric Nash equilibrium. Both requester and provider agents use MARGO, during negotiation, in order to decide their offers and whether to concede or stand still.

[5] http://www.argugrid.eu

```
decision([s(X)]).
decisionrule(r11(X), expensive,
                [s(X), price(X,high), reply(accept)]).
decisionrule(r12(X), cheap,
                [s(X), price(X,low), reply(accept)]).
decisionrule(r21(X), good,
                [s(X), resolution(X,high), reply(accept)]).
decisionrule(r22(X), bad,
                [s(X), resolution(X,low),  reply(accept)]).
decisionrule(r31(X), fast,
                [s(X), deliveryTime(X,low), reply(accept)]).
decisionrule(r32(X), slow,
                [s(X), deliveryTime(X,high), reply(accept)]).
epistemicrule(f11, price(a,high), []).
epistemicrule(f12, resolution(a,low), []).
epistemicrule(f13, deliveryTime(a,high), []).
epistemicrule(f21, price(b,high), []).
epistemicrule(f22, resolution(b,high), []).
epistemicrule(f23, deliveryTime(b,high), []).
epistemicrule(f31, price(c,high), []).
epistemicrule(f32, resolution(c,low), []).
epistemicrule(f33, deliveryTime(c,low), []).
epistemicrule(f41, price(d,low), []).
epistemicrule(f42, resolution(d,low), []).
epistemicrule(f43, deliveryTime(d,low), []).
presumable(reply(accept)).
presumable(reply(reject)).
priority(fast, cheap).
priority(fast, good).
rv([fast]).
sincompatibility(fast,slow).
sincompatibility(cheap,expensive).
sincompatibility(good,bad).
```

Fig. 3. The decision problem of the buyer in the MARGO syntax

5. Related works

Unlike the theoretical reasoning, practical reasoning is not only about whether some beliefs are true, but also about whether some actions should or should not be performed. The practical reasoning (Raz, 1978) follows three main steps: i) *deliberation*, i.e. the generation of goals; ii) *means-end reasoning*, i.e. the generation of plans; iii) *decision-making*, i.e. the selection of plans that will be performed to reach the selected goals.

Argumentation has been put forward as a promising approach to support decision making (Fox & Parsons, 1997). While influence diagrams and belief networks (Oliver & Smith, 1988) require that all the factors relevant for a decision are identified *a priori*, arguments are defeasible or reinstantiated in the light of new information not previously available.

Amgoud & Prade (2009) present a general and abstract argumentation framework for multi-criteria decision making which captures the mental states (goals, beliefs and preferences) of the decision makers. For this purpose, the arguments prescribe actions to reach goals if theses actions are feasible under certain circumstances. These arguments, eventually conflicting, are balanced according to their strengths. Our specific and concrete argumentation framework is in conformance with this approach. The argumentation-based decision making process envisaged by (Amgoud & Prade, 2009) is split in different steps where the arguments are successively constructed, weighted, confronted and evaluated. By contrast, our computation interleaves the construction of arguments, the construction of counterarguments, the evaluation of the generated arguments and the determination of concessions. Moreover, our argumentation-based decision process suggests some decisions even if low-ranked goals cannot be reached.

Bench-Capon & Prakken (2006) formalize defeasible argumentation for practical reasoning. As in (Amgoud & Prade, 2009), they select the best course of actions by confronting and evaluating arguments. Bench-Capon & Prakken focus on the abductive nature of practical reasoning which is directly modelled within in our framework.

(Kakas & Moraitis, 2003) propose an argumentation-based framework for decision making of autonomous agents. For this purpose, the knowledge of the agent is split and localized in different modules representing different capabilities. As (Bench-Capon & Prakken, 2006) and (Amgoud & Prade, 2009), their framework is a particular instantiation of the abstract argumentation (Dung, 1995). Whereas Kakas & Moraitis (2003) is committed to one argumentation semantics, we can deploy our framework to several semantics by relying on assumption-based argumentation.

Rahwan et al. (2003) distinguish different approaches for automated negotiation, including game-theoretic approaches (e.g Rosenschein & Zlotkin (1994)), heuristic-based approaches (e.g. Faratin et al. (1998)) and argumentation-based approaches (e.g. Amgoud et al. (2007); Bench-Capon & Prakken (2006); Kakas & Moraitis (2003)) which allow for more sophisticated forms of interaction. By adopting the argumentation-based approach of negotiation, agents deal naturally with new information in order to mutually influence their behaviors. Indeed, the two first approaches do not allow agents for exchanging opinions about offers. By arguing (even if it is internally), agents can take into account the information given by its interlocutors in a negotiation process (eg. rejecting some offers). Moreover, the agents can make some concessions. In this perspective, Amgoud et al. (2007) propose a general framework for argumentation-based negotiation. They define formally the notions of concession, compromise and optimal solution. Our argumentation-based mechanism for decision making can be used for exploiting such a feature. Morge & Mancarella (2010) have proposed a realisation of the minimal concession strategy. By contrast, we have focus in this paper on the concession-based mechanism of MARGO which model the intuition that high-ranked goals are preferred to low-ranked goals which can be withdrawn. We adopt a decision principle that is so higher-ranked goals should be pursued at the expense of lower-ranked goals, and thus concessions enforcing higher-ranked goals should be preferred to those enforcing lower-ranked goals. To our best knowledge, our argumentation framework is the first concrete system including such a mechanism.

Finally, to the best of our knowledge, few implementation of argumentation over actions exist. CaSAPI[6] (Gartner & Toni, 2007) and DeLP[7] (García & Simari, 2004) are restricted to

[6] `http://www.doc.ic.ac.uk/~dg00/casapi.html`
[7] `http://lidia.cs.uns.edu.ar/DeLP`

the theoretical reasoning. PARMENIDES[8] (Atkinson et al., 2006) is a software to structure the debate over actions by adopting a particular argumentation scheme. GORGIAS[9] (Demetriou & Kakas, 2003) implements an argumentation-based framework to support the decision making of an agent within a modular architecture. Like the latter, our implementation, called MARGO, incorporates abduction on missing information. Moreover, we can easily extend it to compute the competing semantics since MARGO is built upon CaSAPI which is an argumentation engine that implements the dispute derivations described in (Dung et al., 2007).

6. Discussion

To our best knowledge, our argumentation-based mechanism for decision-making is the only concrete argumentation system allowing concessions which is a crucial feature for negotiations. Our framework is built upon assumption-based argumentation frameworks, and provides mechanisms to evaluate decisions, to suggest decisions, and to interactively explain in an intelligible way the choice which has been made to make a certain decision, along with the concessions, if any, made to support this choice. The underlying language in which all the components of a decision problem are represented is a logic-based language, in which preferences can be attached to goals. In our framework, arguments are defined by means of tree-structures, thus facilitating their intelligibility. The concession-based mechanism is a crucial feature of our framework required in different applications such as service selection or agent-based negotiation. Our framework has been implemented and actually exploited in different application domains, such as agent-based negotiation (Bromuri et al., 2009; Morge & Mancarella, 2010), service-oriented agents (Guo et al., 2009), resource allocation (Morge et al., 2009), computational model of trust (Matt et al., 2010) or embodied conversational agents (Morge et al., 2010).
Our decision model only allows qualitative representation of goals. However, in many practical applications, it is not natural to give a quantitative representation of goals. For this purpose, it would be best to have a hybrid approach combining both quantitative and qualitative aspects. Argumentation provides a natural framework for these hybrid systems by providing a link between qualitative objectives and its quantitative representation.

7. Appendix A. Proofs

This appendix includes the proofs considered in this paper.

Proof 1 (Mapping between arguments)**.** *Let* $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ *be a decision framework,* $\mathtt{G} \in \mathcal{G}$ *a set of goals and* $\mathtt{PABFS}_{\mathtt{DF}}(\mathtt{G})$ *be the set of PABFs associated with the goals* $\mathtt{G} \supseteq \mathcal{RV}$*.*

1. *Let* $\bar{A}$ *be a structured argument concluding* $\alpha \in \mathcal{DL}$*. The corresponding argument in some of the PABFs (denoted* $\ell(\bar{A})$*) is defined in Definition 22.*
2. *Let us consider* α *an atomic formula in* $\mathcal{DL}$ *and* $\mathtt{a}$*:* $\mathtt{A} \vdash \alpha$ *be an argument within one PABF in* $\mathtt{PABFS}_{\mathtt{DF}}(\mathtt{L})$*.*
   - *Either* $\alpha \in \mathcal{A}sm_{\mathtt{DF}}$*. Therefore,* α *is deduced by the singleton* $\{\alpha\}$ *and* $\ell^{-1}(\mathtt{A}) = \bar{A}$ *is a hypothetical argument defined such that:*

[8] http://cgi.csc.liv.ac.uk/~katie/Parmenides.html
[9] http://www.cs.ucy.ac.cy/~nkd/gorgias/

$$\begin{aligned}
&\mathtt{conc}(\bar{A}) = \alpha,\\
&\mathtt{premise}(\bar{A}) = \emptyset,\\
&\mathtt{psm}(\bar{A}) = \{\alpha\},\\
&\mathtt{sent}(\bar{A}) = \{\alpha\}.
\end{aligned}$$

- *Or $\alpha \notin \mathcal{A}sm_{DF}$. Therefore, α is deduced by the set of premises $\mathtt{A}$ which, by definition, are all assumptions in $\mathcal{A}sm_{DF}$ and, by construction, there is only one rule r with $\sim \mathtt{deleted}(r) \in \mathtt{A}$ such as $\mathtt{head}(r) = \alpha$.*
  - *Either r is a fact of $\mathcal{T}$ deducing α and so, $\ell^{-1}(\mathtt{A}) = \bar{A}$ is a trivial argument defined such that:*

$$\begin{aligned}
&\mathtt{conc}(\bar{A}) = \alpha,\\
&\mathtt{premise}(\bar{A}) = \{\top\},\\
&\mathtt{psm}(\bar{A}) = \emptyset,\\
&\mathtt{sent}(\bar{A}) = \{\alpha\}.
\end{aligned}$$

  - *Otherwise r is a rule of $\mathcal{T}$ deducing α with a non-empty body and so, $\ell^{-1}(\mathtt{A}) = \bar{A}$ is a tree argument built upon the subarguments $\mathtt{sbarg}(\bar{A})$ defined such that:*

$$\begin{aligned}
&\mathtt{conc}(\bar{A}) = \alpha,\\
&\mathtt{premise}(\bar{A}) = \mathtt{body}(r),\\
&\mathtt{psm}(\bar{A}) = \cup_{\bar{A}' \in \mathtt{sbarg}(\bar{A})} \mathtt{psm}(\bar{A}'),\\
&\mathtt{sent}(\bar{A}) = \mathtt{body}(r) \cup \{\mathtt{head}(r)\} \cup \bigcup_{\bar{A}' \in \mathtt{sbarg}(\bar{A})} \mathtt{sent}(\bar{A}').
\end{aligned}$$

Proof 2 (Mapping between semantics). *Let $\mathtt{DF} = \langle \mathcal{DL}, \mathcal{P}sm, \mathcal{I}, \mathcal{T}, \mathcal{P}, \mathcal{RV} \rangle$ be a decision framework and $\mathtt{AF} = \langle \mathcal{A}(\mathtt{DF}), \mathtt{defeats} \rangle$ be our argumentation framework for decision making. Let us consider $\mathtt{G} \in \mathcal{G}$ with $\mathtt{G} \supseteq \mathcal{RV}$.*

- *Let us consider a s-admissible set of structured arguments $\bar{S}_1$ concluding $\mathtt{G}_1 \subseteq \mathtt{G}$. Due to the lemma 1, we can built the set of arguments $\mathtt{S}_1$ such that for any structured argument $\bar{A}_1 \in \bar{S}_1$ there is an argument $\mathtt{a}_1 \in \mathtt{S}_1$, where $\mathtt{a}_1 : \ell(\bar{A}_1) \vdash L$ is in some $\mathtt{PABFS}_{DF}(\mathtt{G})$. We consider here $\mathtt{pabf}^1_{DF}(\mathtt{G}_1) \in \mathtt{PABFS}_{DF}(\mathtt{G})$ where all the arguments appear. Due to the construction of $\mathtt{S}_1$, the set of arguments $\mathcal{A}(\ell(\bar{S}_1))$ is conflict-free and defend itself within $\mathtt{pabf}^1_{DF}(G)$. Therefore, $\mathtt{S}_1$ is an admissible set. Let us consider a different $\mathtt{pabf}^2_{DF}(\mathtt{G}_2) \in \mathtt{PABFS}_{DF}(\mathtt{G})$ such that $\mathtt{pabf}^2_{DF}(\mathtt{G}_2)\mathcal{P}\mathtt{pabf}^1_{DF}(\mathtt{G}_1)$. Due to the definition 21, $\mathtt{G}_2 \supseteq \mathtt{G}_1$ and $\forall g_2 \in \mathtt{G}_2 \setminus \mathtt{G}_1$ there is no $g_1 \in \mathtt{G}_1$ such that $g_1 \mathcal{P} g_2$. If we suppose that $\mathtt{pabf}^2_{DF}(\mathcal{G})$ contains an admissible set of arguments deducing $\mathtt{G}_2$, then the corresponding set of structured arguments concluding $\mathtt{G}_2$ is admissible. It is not the case.*
- *Let us consider $\mathtt{pabf}_1 \in \mathtt{PABFS}_{DF}(\mathtt{G})$ which contains an admissible set of assumptions $\mathtt{A}_1$ deducing $\mathtt{G}_1$ with $\mathcal{RV} \subseteq \mathtt{G}_1 \subseteq \mathtt{G}$. If we suppose that there is no $\mathtt{pabf}_2 \in \mathtt{PABFS}_{DF}(\mathtt{G})$, with $\mathtt{pabf}_2 \mathcal{P} \mathtt{pabf}_1$, which contains an admissible set of assumptions deducing $\mathtt{G}_2 \subseteq \mathtt{G}$ with $\mathtt{G}_2 \mathcal{P} \mathtt{G}_1$, then the corresponding s-admissible set of structured arguments $\ell^{-1}(\mathtt{A}_1)$ concludes $\mathtt{G}_1$ and there is no other s-admissible set of structured arguments $\bar{S}_2$ concluding $\mathtt{G}_2$.*

8. Acknowledgments

We would like to thank the anonymous reviewers for their detailed comments on this paper. The authors thank Phan Minh Dung and Francesca Toni for many useful discussions on the topic of this work.

9. References

Amgoud, L. & Cayrol, C. (1998). On the acceptability of arguments in preference-based argumentation, *Proc. of UAI*, Morgan Kaufmann, Madison, Wisconsin, USA., pp. 1–7.

Amgoud, L. & Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments, *Annals of Maths and AI* 34(1-3): 197–215.

Amgoud, L., Dimopoulos, Y. & Moraitis, P. (2007). A unified and general framework for argumentation-based negotiation, *Proc. of AAMAS*, Honolulu, Hawaii, pp. 963–970.

Amgoud, L. & Prade, H. (2009). Using arguments for making and explaining decisions, *Artificial Intelligence Journal* 173(3-4): 413–436.

Atkinson, K., Bench-Capon, T. & McBurney, P. (2006). Computational representation of practical argument, *Synthese, special issue on Knowledge, Rationality and Action* 152(2): 157–206.

Barber, S., Bossert, W. & Attanaik, P. K. (2004). *Handbook of Utility Theory*, Vol. 2, Kluwer Academic Publisher, chapter Ranking Sets of Objects, pp. 893–977.

Bench-Capon, T. (2002). Value based argumentation frameworks, *Proc. of the 2nd International Workshop on Non-Monotonic Reasoning (NMR)*, pp. 444–453.

Bench-Capon, T. & Prakken, H. (2006). Justifying actions by accruing arguments, *Proc. of the 1st International Conference on Computational Models of Argument*, IOS Press, pp. 247–258.

Bondarenko, A., Toni, F. & Kowalski, R. (1993). An assumption-based framework for non-monotonic reasoning, *Proc. of LPNMR*, MIT Press.

Bromuri, S., Urovi, V., Morge, M., Toni, F. & Stathis, K. (2009). A multi-agent system for service discovery, selection and negotiation (demonstration), *8th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009), Budapest, Hungary, May 10-15, 2009, Volume 2*, IFAAMAS, pp. 1395–1396.

Chesñevar, C. I., Maguitman, A. G. & Loui, R. P. (2000). Logical models of argument, *ACM Comput. Surv.* 32(4): 337–383.

Demetriou, N. & Kakas, A. C. (2003). Argumentation with abduction, *Proc. of the 4th Panhellenic Symposium on Logic*.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intell.* 77(2): 321–357.

Dung, P. M., Kowalski, R. A. & Toni, F. (2006). Dialectic proof procedures for assumption-based, admissible argumentation, *Artif. Intell.* 170(2): 114–159.

Dung, P. M., Mancarella, P. & Toni, F. (2007). Computing ideal sceptical argumentation, *Artificial Intelligence, Special Issue on Argumentation* 171(10-15): 642–674.

Faratin, P., Sierra, C. & Jennings, N. R. (1998). Negotiation decision functions for autonomous agents, *International Journal of Robotics and Autonomous Systems* 24: 3–4.

Fox, J. & Parsons, S. (1997). Argumentation and decision making, *Proc. of IEE Colliquium on Decision Making and Problem Solving*, London.

García, A. J. & Simari, G. R. (2004). Defeasible logic programming: An argumentative approach, *Theory and Practice of Logic Programming* 4(1-2): 95–138.

Gartner, D. & Toni, F. (2007). CaSAPI: a system for credulous and sceptical argumentation, *Proc. of ArgNMR*, pp. 80–95.

Guo, L., Ghanem, M., Curcin, V. & Azam, N. (2009). Intelligent agents in the service-oriented world - an industrial experience report, *Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on* 1: 693–696.

Jennings, N. R., Faratin, P., Lomuscio, A. R., Parsons, S., Sierra, C. & Wooldridge, M. (2001). Automated negotiation: prospects, methods and challenges, *International Journal of Group Decision and Negotiation* 10(2): 199–215.

Kakas, A. & Moraitis, P. (2003). Argumentative-based decision-making for autonomous agents, *Proc. of AAMAS*, ACM Press, pp. 883–890.

Kowalski, R. & Toni, F. (1996). Abstract argumentation, *Artificial Intelligence and Law Journal Special Issue on Logical Models of Argumentation* 4(3-4): 275–296.

Matt, P.-A., Morge, M. & Toni, F. (2010). Combining statistics and arguments to compute trust, *in* W. van der Hoek, G. Kaminka, Y. LespÃl'rance, M. Luck & S. Sen (eds), *Proceedings of the ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'2010)*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 209–210.

Morge, M., Abdel-Naby, S. & Beaufils, B. (2010). Towards a dialectical approach for conversational agents in selling situations, *in* I. R. Peter McBurney & S. Parsons (eds), *The Seventh Proceedings of the seventh International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'2010)*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 127–144.

Morge, M. & Mancarella, P. (2010). Assumption-based argumentation for the minimal concession strategy, *in* P. McBurney, I. Rahwan, S. Parsons & N. Maudet (eds), *Argumentation in Multi-Agent Systems, 6th International Workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009. Revised Selected and Invited Papers*, Vol. 6057 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Germany, pp. 114–133.
URL: *http://www.springerlink.com/content/902635985hw042n8/*

Morge, M., Mathieu, P. & Mancarella, P. (2009). Assumption-based argumentation for the minimal concession strategy of agents engaged in resource negotiation, *Actes des CinquiÃÍmes JournÃl'es Francophones sur les ModÃÍles Formels de l'Interaction (MFI'09)*, pp. 232–242.

Oliver, R. M. & Smith, J. Q. (eds) (1988). *Influence Diagrams, Belief Nets and Decision Analysis*, John Wiley and Sons.

Prakken, H. & Sartor, G. (1997). Argument-based logic programming with defeasible priorities., *Journal of Applied Non-classical Logics* 7: 25–75.

Rahwan, I., Ramchurn, S. D., Jennings, N. R., McBurney, P., Parsons, S. & Sonenberg, L. (2003). Argumentation-based negotiation, *The Knowledge Engineering Review* 18(4): 343–375.

Raz, J. (ed.) (1978). *Practical Reasoning*, Oxford University Press, Oxford, UK.

Rosenschein, J. S. & Zlotkin, G. (1994). *Rules of encounter: designing conventions for automated negotiation among Computers*, The MIT press series of artificial intelligence, MIT press.

Simari, G. R. & Loui, R. P. (1992). A mathematical treatment of defeasible reasoning and its implementation, *Artif. Intell.* 53(2-3): 125–157.

Stournaras, T. (ed.) (2007). *Concrete scenarios identification & simple use cases*, Deliverable document D1.1 ARGUGRID.

Toni, F., Grammatikou, M., Kafetzoglou, S., Lymberopoulos, L., Papavassileiou, S., Gaertner, D., Morge, M., Bromuri, S., McGinnis, J., Stathis, K., Curcin, V., Ghanem, M. & Guo, L.

(2008). The argugrid platform: An overview, *Proc. of the 5th International Workshop on Grid Economics and Business Models (GECON'08)*, Vol. 5206 of *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Las Palmas de Gran Canaria, Spain, pp. 217–225.

Vreeswijk, G. (1997). Abstract argumentation systems, *Artif. Intell.* 90(1-2): 225–279.

13

Academic Orientation Supported by Hybrid Intelligent Decision Support System

Emilio J. Castellano, Manuel J. Barranco and Luis Martínez
Department of Computer Sciences/University of Jaén
Spain

1. Introduction

It is common that in all academic systems the students must make decisions about the future by choosing among different alternatives that include professional profiles or *modalities*, elective subjects or *optional*, etc. These decisions can have a very important influence in the academic journey of the students because sometimes a wrong decision can lead to academic failure or its correction implies a time cost for the students.

It is remarkable that in many academic systems these important decisions have to be made in early stages in which the students do not have enough maturity or knowledge to be conscious about the consequences in their future if a wrong decision is made. Ideally these multiple choices, offered to the students, want to facilitate the acquirement of some professional and valuable competences to obtain a job. Taking into account that the suitability of people in jobs or studies is not only restricted to their taste or preferences, but also other factors involved in the process of acquiring maturity as an adult who develops any function. These factors such as capacities, skills, attitudes concerning the task, social attitudes, taste, preferences, etc. (Briñol 2007, Robertson 1993, Salgado 1996, Shevin 2004), must be taken into account in such processes.

Initially these decision making processes were made by the students themselves or with their parents support according to different criteria such as, preferences, future job market, even randomly, etc. Therefore, in order to improve this situation different countries introduced one figure, so-called *advisor*, whose role is to guide the students in their decision making situations regarding their academic future.

The academic orientation process carried out by these advisors imply the review of different information regarding the students in order to report which academic alternatives suits better their skills, competences and needs. In most of academic institutions the advisor deals yearly with several hundreds of students with different skills, personalities, etc. To make an idea, in Spain, and depending on the high school, advisors can manage from 200 to 800 students. This number of students implies a big search space to find the relevant information that can facilitate the orientation for each one in a right way, making very hard to perform successfully the academic orientation process. Hence it seems suitable the development of automated tools that support the accomplishment of the different processes of academic orientation to improve the success of advisors' tasks.

An overview of different problems in the real world that deal with search spaces drove us to pay attention to the situation raised some years ago with the advent of Internet and the

increase of requirements in many areas. One of the most demanded solutions is based on the necessity of finding out suitable items over huge and/or complex search spaces in *e-shops*. The users need help to explore and filter all the possibilities about the items offered in order to improve the quality of their choices, minimize the time consumed and the wrong decisions.

Different tools have been developed to accomplish the previous goals, being remarkable the use of Recommender Systems (Adomavicius 2005, Resnick 1997, Rodríguez 2010). These systems offer recommendations to users according to their preference profiles, guiding them through search spaces in order to find out the most suitable items for their needs in many real-life situations. The growth of this area is basically due to the vast amount of available information in Internet and the facilities provided by the Web to create users' communities in order to share experiences, tastes, interests, etc. In this sense, Recommender Systems (RS) provide users customized advices and information about products or services of interest in order to support their decision-making processes. Usually a RS estimates the degree of like or dislike either how suitable is an item for a user. To do so, a user profile is generated from his/her preferences which are built by gathering explicit data, implicit data or both (Adomavicius 2005, Pazzani 1999) (see Figure 1).

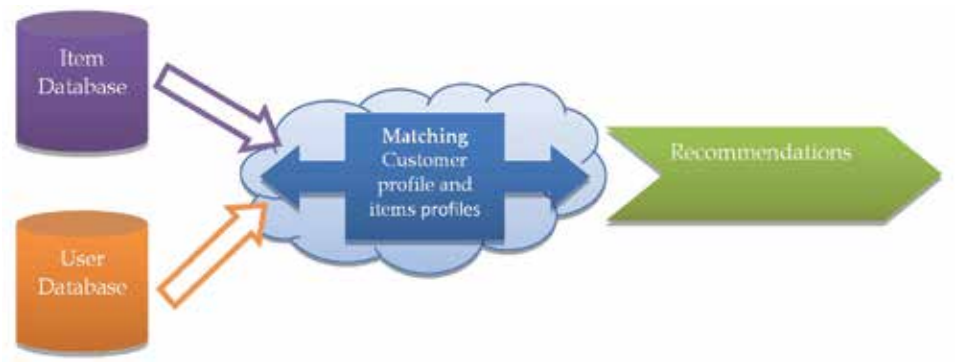


Fig. 1. Recommendation scheme

Recommendations suggested by RS can be obtained in different ways (Resnick 1997, Schafer 2001) and hence there exist different types of RS, depending on the information and the technique utilized to compute its recommendations:

- *Content-based Recommender System* (Horvath 2009, Martínez 2007b): These RS are based on features of items experienced by the customer in order to create users' profiles and use these to find out similar items.
- *Collaborative Filtering Recommender Systems* (Goldberg 1992, Takacs 2009): Instead of building users' profiles based on item features, they use customers' ratings to compute a recommendation of items for a specific user considering the similarity of the target customer and the rest of users.
- *Knowledge Based Recommender Systems* (Burke 2000, Zhen 2010): These systems base their recommendations on inferences about user's needs and preferences. They use the knowledge about customers' necessities and how an item matches these necessities to infer useful recommendations.
- *Demographic Recommender Systems* (Krulwich 1997): They categorize customers into groups using stereotype reasoning based on the information stored in the user profile that contains different demographic features.

- *Utility Based Recommender Systems* (Guttman 1998) compute their recommendations on the match between the user's need and the available options that could maximize the accomplishment of the expected utility.
- *Hybrid Recommender Systems* (Albadvi 2009, Burke 2002b): To overcome the disadvantages of previously mentioned recommender systems sometimes is used hybrid techniques which try to make useful the best of methods used in hybridization.

The most spread recommender systems are collaborative and content based systems both have provided good results in different areas as tourism (Sebastia 2009), e-learning (Romero 2009), academic orientation (Castellano 2009b), etc.

In (Castellano 2009b) was introduced the use of collaborative filtering for academic orientation, in the case of study of Spanish Academic System, by using *marks* as ratings of user's profile that are filtered in order to support advisors in their orientation for the students. In (Castellano 2009a) these processes were merged with fuzzy techniques (Martínez 2008, Martínez 2007a, Zadeh 1965) to improve the comprehension of the results by the advisors.

Even though the DSS, *OrieB*, presented in (Castellano 2009a, Castellano 2009b) provides good results in general, it suffers the same weaknesses and limitations that any collaborative RS (Herlocker 2004, Lee 2003, Pazzani 1999), such as, *grey sheep, historical data, cold-start, scarcity*, etc. Therefore, we detected a problem in *OrieB*, because academic systems are not static but they suffer changes quite often and new subjects can appear due to modification or adaptation of academic profiles. Hence these changes imply the appearance of *cold start* and *historical data* problems decreasing the performance of the system for academic orientation.

In this chapter we aim to improve the performance of academic orientation support by using a hybrid technique that supplies support even in those situations in which there is scarcity information or new items. First, we should find what type of information would be available when marks do not exist yet for new subjects and choose the techniques that should be hybridizing with collaborative filtering to obtain good results in the academic orientation situations in which such a filtering technique is not enough. In this case the *Competency based Education* (CBE), that is an emerging curriculum model that tries to satisfy the demands of learning contexts, by the developing competencies, enabling students to act in a complex world in constant transformation (Zalba 2006), might play an important role. A competence is the ability to perform effectively in a given situation, it is based on knowledge but it is not limited to it (Perrenoud 1999). Competences are complex knowledge and represent the know-how to integrate conceptual, procedural and attitudinal knowledge. The current importance and relevance of the CBE is showed in Tuning Educational Structures in Europe (Europe TUNING 2000). As well in different countries this educational paradigm is being developed in secondary education and high schools and students are now being evaluated focusing on competences. So the use of *competences* will be very useful in our aims because they keep relevant information for academic orientation.

Therefore, once we know the information available when the collaborative filtering is not working our second step is to define a hybrid model that that hybridizes *collaborative* and *content-based* techniques (CB), where the content-based model will be based on the textual description of subject competences. This model will be implemented in a DSS for Academic Orientation. In such a way the DSS can provide more and better support to the academic advisors in their task.

This chapter is organized as follows: Section 2 reviews recommender systems; Section 3 introduces academic orientation and competency based education; Section 4 shows the use of collaborative filtering in academic orientation and points out its weaknesses; Section 5 proposes a hybrid model for academic orientation based on collaborative and content-based techniques; Section 6 presents a Decision Support System that incorporates such a model and Section 7 concludes this chapter.

2. Recommender systems

As we have pointed out previously the techniques that we will use to support academic orientation will be based on those ones used in recommender systems that support customers in their buying processes in the e-commerce arena where customers face to huge amounts of information about items that are hard to check in an affordable time in order to buy the most suitable item/s.

Notwithstanding, there exist several techniques in the recommender systems, in this section we will only focus on the revision of collaborative, content-based and hybrid ones because they will be the used in our proposal.

2.1 Collaborative recommender systems

Collaborative recommender systems (CRS) collect human opinions of items, represented as ratings, in a given domain and group customers with similar needs, preferences, tastes, etc., in order to recommend active user items which liked in the past to users of the active user group (Herlocker 1999).

Most of CRS use explicit data directly provided by users and related to customers' perceptions and preferences that implies uncertainty, though it has been fairly usual the use of precise scales to gather such information, the use of linguistic information to model such an information seems more suitable and several proposals have been developed (Martínez 2007b, Porcel 2009).

There exist different collaborative approaches (Adomavicius 2005): (i) *Memory-based* which use heuristics that predict ratings basing on the whole dataset of previously rated items and (ii) *Model-based* which use the collection of ratings to learn a model capable of predicting ratings. According to Figure 2, both models fulfill three general tasks to elaborate the recommendations demanded by users:

- *Analyzing and selecting data sets*: data from ratings must be collected and optimized for the system (Herlocker 2004).
- *Grouping users* with similar tastes and preferences in order to compute recommendations basing on a similarity measure as Pearson Correlation Coefficient (Adomavicius 2005, Breese 1998).
- *Generating predictions*: Once users have been grouped by interest (similarity), the system uses them to compute predictions for the target customer by using different aggregation methods (Breese 1998, Herlocker 1999).

Collaborative filtering methods provide several advantages regarding other techniques used in recommender systems (Herlocker 1999, Sarwar 2001):

- Capability to manage information whose content is not easily analyzed automatically, because they do not need knowledge domain, i.e., no information or knowledge about the products is needed.

- Ability to filter items based on quality and taste, not only on its features.
- Ability to provide serendipitous recommendations. Other systems never recommend products which are *outside the box*, i.e., recommended products are not very different to the ones positively rated by the customer.
- Adaptability, its quality is improved along the time. When the number of customers and rates increases these systems work better.

Fig. 2. Processes performed in the collaborative recommendation scheme

Despite the general good performance of these systems, they present several weaknesses and limitations:

- *The cold-start problem*: This problem is presented with both users and products. When a new user access to the system, it has not any information about him/her. Therefore, the system cannot compare him/her with the users of the database and cannot provide recommendations. When a new item is added and it has not been assessed by any user yet, it cannot be recommended. This is the problem we focus on this paper, so it will be further detailed below.
- *The "grey sheep" problem:* This system does not work properly with "grey sheep" users, which are in the frontier of two groups of users.
- *Historical data set:* Being an adaptive system can be an advantage but can be also a disadvantage when the system is starting and the historical data set is small.

Obviously, no recommender system can work without some initial information but the quality and efficiency of the system depends on the ability to predict successful recommendations with the minimum amount of information about users and items. Some solutions to this problem require knowledge about the items, or content based information, for example, a movie recommender system needs to know attributes like actors, the genre, etc. This kind of knowledge is not always available or is scarce. Other proposed solutions are only partial solutions because improve the recommendations when the data about the user is small but do not work when this set is empty (new user).

2.2 Content-based recommender systems

Content-Based Recommender Systems (CBRS) recommend similar items to those ones that the user liked in the past by comparing various candidate items with items previously rated by the user and the bestmatching item(s) are recommended (Adomavicius 2005, Jung 2004, Symeonidis 2007). The content-based approach to recommendation has its roots in information retrieval and information filtering research (Adomavicius 2005, Martínez 2007a, Pazzani 1999). CBRS need user profiles that contain information about users' tastes, preferences, and needs. Profiles can be built using information collected from users

explicitly, e.g., through questionnaires, or implicitly inferred from user behavior over time. On the contrary to collaborative recommender systems, the content based ones assume that the user does not interact with other users. These systems exploit the information related to the description content of the previously evaluated items such that the system learns the relationship between a single user and the description of the new items. Let S be a set of items to be recommended, the description of each item s_i is provided by m features c_j, such as is showed in Table 1.

	c_1	…	c_j	…	c_m
s_1	v_{11}	…	v_{1j}	…	v_{1m}
…	…	…	…	…	…
s_n	v_{n1}	…	v_{nj}	…	v_{nm}

Table 1. Item description

In content-based recommendation methods each user's profile is built with information related to previous items selected by the user in the past. Therefore the user's profile P of user u is computed based on the description of previously experienced items and optionally some explicit rating r_i about them *(Adomavicius 2005)*.

	c_1	…	c_m	R_u
s_1^u	v_{11}^u	…	v_{1m}^u	r_1^u
…	…	…	…	…
$s_{n_u}^u$	$v_{n_u 1}^u$	…	$v_{n_u m}^u$	$r_{n_u}^u$
P_u	p_1^u	…	p_m^u	
W_u	w_1^u	…	w_m^u	

Table 2. User Profile

The user's profile (see Table 2) is used to determine the appropriateness of the item for recommendation purposes by matching it with the items descriptions (see Table 1). Since, as mentioned earlier, CBRS are designed mostly to recommend text-based items, the content in these systems is usually described with keywords which can represent, for example, features of an item.

More formally, let *ContentBasedProfile(c)* be the profile of user, c, containing tastes and preferences of this user, for example a vector of weights (w_{c1}, …, w_{ck}) where each weight w_{ci} denotes the importance of keyword k_i to user c.

In content-based systems, the utility function $u(c,s)$ is usually defined as:

$$u(c,s) = score(ContentBasedProfile(c), Content(s)) \tag{1}$$

Both *ContentBasedProfile(c)* of user c and *Content(s)* of document s can be represented as vectors $\vec{v_c}$ and $\vec{v_s}$ of keyword weights. Moreover, utility function $u(c,s)$ is usually represented in the information retrieval literature by some scoring heuristic defined in terms of vectors $\vec{v_c}$ and $\vec{v_s}$, such as the cosine similarity measure

$$u(c,s) = \cos\left(\vec{v_c}, \vec{v_s}\right) = \frac{\vec{v_c} \bullet \vec{v_s}}{\vec{v_c}_2 \times \vec{v_s}_2} = \frac{\sum_{i=1}^{K} v_{i,c} v_{i,s}}{\sqrt{\sum_{i=1}^{K} v_{i,c}^2} \sqrt{\sum_{i=1}^{K} v_{i,s}^2}} \tag{2}$$

where K is the total number of keywords in the system. Depending on the information and the content based technique used, such keywords or features could be either equally important or weighted according to their relevance, for example, using TF-IDF (frequency/inverse document frequency) which is one of the best-known measures for specifying keyword weights in Information Retrieval (Adomavicius 2005, Symeonidis 2007).
Although CB approach enables personalized and effective recommendations for particular users, it has also some disadvantages (Adomavicius 2005, Lenar 2007, Symeonidis 2007):

- Limited Content Analysis: Content-based techniques are limited by the features that are explicitly associated with the objects that these systems recommend. Therefore, in order to have a sufficient set of features, the content must either be in a form that can be parsed automatically by a computer (e.g., text) or the features should be assigned to items manually. Another problem with limited content analysis is that, if two different items are represented by the same set of features, they are indistinguishable.
- Overspecialization: When the system can only recommend items that score highly against a user's profile, the user is limited to being recommended items that are similar to those already rated. In certain cases, items should not be recommended if they are too similar to something the user has already seen.
- CB is based only on the particular user relevance evaluations, but users usually are very reluctant to give them explicit, so usually other implicit, possibly less adequate, methods must be used.

To overcome these problems, CB and CF techniques have been combined to improve the recommendation procedure.

2.3 Hybrid recommender systems

Due to limitations observed in both previous recommendation techniques, nowadays it has been used to overcome these limitations the hybridization technique. This technique combines two or more recommendations methods in order to obtain a better performance and/or accuracy that in each of the methods separately. It is common to combine collaborative filtering technique (CF) with another method to avoid cold start problem.
Following it is presented the classification of hybridizing techniques for recommender systems presented by Burke (Burke 2002a):

1. *Weighted* systems: recommendations of each system are combined giving each one a specific weight of the final recommendation, depending on the system which computed them. Importance of each item is computed based on results obtained from recommendation techniques present in the system. All capacities of the system are used in recommendation process in an easy and direct way, and also it is easy to adjust weights manually or by simple algorithms. However, this technique start from the hypothesis of giving a uniform importance value to each of the distinct techniques composing the system, and this is not always true.
2. *Switching* hybrid systems use several criteria to alternate recommendation technique whenever is needed each moment. This method brings an additional complexity in the recommendation process as it is needed to determine choosing method criteria, as it is another parameterization level. On the other hand, if the criterion is selected properly it can take advantage of qualities of composing recommendation systems and avoid weakness of systems in those situations.

3. *Mixed* hybrids result from various recommendation methods showing them at the same time. It can be used in situations where a big number of recommendations are required. So can be avoid the *new item* problem if always content-based recommendations are shown, because this component compute its recommendations using features of new items, regardless they have been rated or not. However, *new user* problem is not solved.
4. *Feature combination*: another way of hybridization is to manage collaborative information as a data feature associated with each example, and to use content-based techniques over this extended dataset. This kind of hybridization allows using collaborative data without an exclusive dependence, decreasing system sensibility to the number of users that have rated an item. Moreover, allow system to obtain information about inherent similarity not possible to find in content-based systems.
5. *Cascade* method combines techniques by means of process composed of several phases: preliminary selection of items to recommend by means of one of the methods. In following phases other techniques are used to filter the set of candidates. This technique avoids second method to manage items discarded by first method, or with enough negative ratings to never be recommended. Thanks to this, phases next to the initial one are more efficient. Moreover, an item rated negatively cannot be recommended because recommendations are refined in each phase.
6. *Feature augmentation*: this technique produces a valuation or a classification of an item as a new keyword or feature, and this information is incorporated in the process using next recommendation method. This method is interesting because it allows improving performance of main recommendation technique without altering original operation and conception.
7. *Meta-level*: another way of combining several recommendation techniques is using model generated by a method as input for another. This differs from *increase of features* as this use model learned to generate features as inputs, not the whole model. The profit of this kind of hybridization is that model apprehended is a compressed representation of user's interests and collaborative filtering operates easier with these datasets.

As we have seen, hybridization can support the overcoming of some problems associated with certain recommendation techniques. Although hybrid recommender systems which use collaborative and content-based methods always will present the *cold start* problem as both need a dataset to operate, once the system has a minimum dataset it can overcome certain limitations inherent to collaboration filtering.

3. Academic background

In section 1 was introduced the aim of this chapter: to develop a DSS for *academic orientation* by using a hybrid model that uses a *competency based education* paradigm, in order to overcome the new subject problem. To facilitate the understanding of the proposal this section reviews some concepts and issues related to educational systems, academic orientation and competency based education, focusing on the Spanish Academic System.

3.1 Educational systems

The concept of academic orientation is related to the student curriculum guidance, it means that students have to make decisions about their curriculum in order to obtain a degree in the topic they prefers the most or their skills are the most appropriate. So the academic

orientation consists of supporting students in such decisions helping them by means of advices and additional information to facilitate their decisions, such that, students will be successful in their academic choice.
In order to make clear the concept of academic orientation we have studied different educational systems to extract common features in order to show the generality of our proposal. We have observed two common features: Evaluation and Specialization.

3.1.1 Evaluation

The main point that all academic institutions and educational systems have in common, is that they evaluate their students by means of different evaluation tools (tests, essays, tasks, exercises, etc.). The final result of this process is a *mark* that reflects not only the students' knowledge but also their skills, preferences, tastes about the subject, etc.
The starting point for our proposal is composed by three information items: students, subjects and marks (see Table 3).

	Mathematics	Literature	Biology	Economy
John	9	6	4	8
Miranda	5	9	Not available	6
Claire	4	3	7	Not available
William	7	2	Not available	6

Table 3. A fragment of a rating/mark matrix for students and subjects

The main point that all academic institutions and educational systems have in common, is that they evaluate their students by means of different evaluation tools (tests, essays, tasks, exercises, etc.). The final result of this process is a *mark* that reflects not only the students' knowledge but also their skills, preferences, tastes about the subject, etc.
The starting point for our proposal is composed by three information items: students, subjects and marks (see Table 3).

3.1.2 Specialization

Most educational systems all over the world from early educational stages to University degrees allow students to choose among different specialization branches according to their skills, preferences, attitudes and marks, building a personalized so-called Academic Profile.
These specialization academic branches are based on certain patterns. Each branch consists of a set of subjects: several ones are compulsory, so-called core subjects, and others, the elective subjects, are optional.
On the other hand, an academic branch can group subjects in different modules which try to specialize students in a topic or area. These modules are called profiles or modalities and may be different depending on each country and sometimes on the institution. The modalities consist of modality and elective subjects. The former are specific of the modality although can be shared by several modalities. The latter can be selected independently of the modality.
Most of academic institutions (Secondary school, High school, Universities) follow this scheme by offering at least core and elective subjects, adding others the possibility of choosing modalities and their modality subjects, in order to build an Academic Profile. For example, in a Computer Engineering degree the student can specialize in software, and

within this area, he or she can choose to be an expert in Recommender Systems, by choosing artificial intelligence, programming and object oriented databases subjects. So, an Academic Profile concerns several subjects of each group.
The point is that, to reach this level of specialization in a specific area in which a student is supposed to be more capable, the student needs to make decisions in order to obtain the appropriate knowledge and abilities. The more accurate those decisions are the better the development of the student's potential.

3.2 Academic orientation tasks. Advisors

Students must make hard decisions about the future since early ages despite their personality and maturity could not be enough to make properly those important decisions. So that, some educational systems have created a figure to guide the students in their academic orientation decision making, so-called Advisor.
Without loss of generality we focus on the Spanish Educational System, advisors for secondary and high schools (other stages are similar).
In Spain the advisor's duties are mainly three:

a. Diversity management
b. Personal attention
c. Academic-professional orientation.

We will focus our paper on the Academic Orientation task. Usually, each advisor deals yearly with several hundreds of students (between 200 and 500 each year), depending on the institution. Therefore, with such a number of students where each one has his/her own personality and skills, the advisor's tasks are really hard to perform successfully for all students. The development of supporting tools for advisors can then improve the success of their tasks.
Regarding Academic Orientation, the advisors usually face two main types of students.

- Students with no idea about what profile to choose. Advisor should help them to build their academic profile by choosing modality and subjects.
- Students that want to study a profile independently of their skills to acquire such a specialization. Here, advisors can help students if they are able to identify topics or subjects in which those students can find difficulties to achieve successful results.

3.3 Competency based education

In a changing world based in changing technology and services, it is necessary for academic systems to have solid foundations capable of allowing a flexible configuration of learning process formulated as flexible academic profiles. Zalba (Zalba 2006) considers that a design and development of a Competency Based Education (CBE) is the best choice in order to set up a curriculum based on social context and necessities.
Actually it is necessary to introduce the concept of academic competence or competence to generalize. It is difficult to define objectively what an Academic Competence is, but it can be understood as a multidimensional construct composed of the skills, attitudes, and behaviors of a learner that contribute to academic success in the classroom. In general, a competence must entail a set of skills of problem solving — enabling the individual to resolve genuine problems or difficulties that he or she encounters and, when appropriate, to create an effective product — and must also entail the potential for finding or creating problems — and thereby laying the groundwork for the acquisition of new knowledge (Gardner 1993).

The CBE is an emerging curriculum model that tries to satisfy the demands of learning contexts, by the developing competencies, enabling students to act in a complex world in constant transformation (Zalba 2006). In short, a competence is the ability to perform effectively in a given situation, it is based on knowledge but it is not limited to it (Perrenoud 1999). Competences are complex knowledge and represent the know-how to integrate conceptual, procedural and attitudinal knowledge.

The current importance and relevance of the CBE is showed in Tuning Educational Structures in Europe. The contribution of universities to the Bologna process and the Tuning Latin America project (González 2008). In this new context, competences serve as well as unification and comparability tools. They are a measure of students' capabilities development, and a start point and guidance not only for subjects but for academic and professional profiles, in which competences emerge as an important element to guide the selection of knowledge appropriate to particular ends. Learners can develop a particular competence or set of competences in order to finally acquire certain knowledge, attitudes and professional capabilities. Tuning serves as a platform for developing reference points at subject area level. These are relevant for making programs of studies (bachelor, master, etc.) comparable, compatible and transparent. Reference points are expressed in terms of learning outcomes and competences. Learning outcomes are statements of what a learner is expected to know, understand and be able to demonstrate after completion of a learning experience. According to Tuning, learning outcomes are expressed in terms of the level of competence to be obtained by the learner.

In contemporary educational systems using CBE, flexible profiles are defined basing on the possibility of choosing with flexibility different subjects from a determined set. So, if students want to create their own academic profile, they only have to choose the subjects they consider interesting for their purposes. But, in order to choose correctly, it is necessary to take into account that subjects are affected by competences in two ways:

- A subject needs that students accomplish certain level of development for specific competences before students are ready to confront this subject.
- Each subject itself contributes to the development of certain competences in a major or minor degree.

So, in order to acquire certain development of competences, learners must study subjects in a retroactive way, so that stepwise they develop the desired competences. In other words, to course certain subjects it is necessary to have studied and passed those subjects capable of develop competences prerequisite of more advanced subjects.

We will take these ideas as basis for our further explanation of the use of CB in Academic Orientation.

4. *OrieB*: Collaborative filtering in academic orientation

In (Castellano 2009b) it was presented *OrieB* a DSS based on collaborative filtering which offers recommendations to students as subjects susceptible of being good elections in order to perform an adequate academic path. As collaborative techniques deal with *customers, items and ratings*, *OrieB* deals with *students, subjects and marks*, so the dataset about Academic Orientation is adapted to apply CF techniques. Another point is that we considered generically that students with similar marks share similar skills. So if we analyze the performance of students in a given group, G_i, in different curriculum modalities. This analysis might be then utilized to support future students classified in G_i for their academic decisions so that *OrieB* consists of:

a. **Dataset**: a dataset with students' marks obtained from coursed subjects.
b. **Grouping students**: students are grouped based on the similarity of its marks and those from other students. This similarity is computed used Pearson Correlation Coefficient (Castellano 2009b).
c. **Predictions**: system makes a prediction about the mark that student would obtain for candidate subjects so this prediction could be used by the advisors in order to support student decisions.

Data analysis about the performance of CF in academic orientation and hence in *OrieB* was done and presented in (Castellano 2009b). The results the survey showed that the system was capable to support successfully *advisors* in an accurate way.

A further detailed explanation about the previous survey, the decision support model and OrieB can be revised in (Castellano 2009b).

This approach has certain limitations suffered by Orieb which we are trying to solve in this chapter:

- *New item problem*: whenever institutions due to either new laws or academic needs a new subject is required, OrieB has no possibility of recommend it as it has no marks available from students, because no one has study this subject yet and no one has a mark, and CF is unable to works without this kind of data. This limitation could be overcome with a content-based recommendation which studies competences required for the subject and student's competences developement.
- *The "grey sheep" problem:* Although OrieB is not much affected for this limitation because the range of elections is much closed, CBRS would offer more personalization because it would take into account not only subject's mark prediction, but also information about that subject.
- Same as before, *historical data set* problem can be solved by using actual student's competences development and information about subjects.

As we can see, the use of a CBRS can help improving OrieB in several ways. Next section will explain how to context a CBRS in Academic Orientation area.

5. CB and academic orientation

So far we have seen the use of CF in academic orientation and some limitations that it presented such the cold start. Consequently a good way to overcome the cold start might be the use of the hybridizing with content based techniques as it has been done in other applications.

In this section we propose a content based model for academic orientation that afterwards it will be hybridized with the CF technique already revised.

It has been pointed out that to perform efficiently a CBRS requires:

- Item profiles which contains information to be used in the recommendation process
- User profiles in order to match item profiles searching adequate candidates.

These profiles need to be based on a set of keywords or features. In our case, items will be subjects and users will be students. Features chosen to elaborate profiles are Academic Competences.

Following European Union guidelines, Spanish Academic System is based in eight competences which are:

- Linguistic communication
- Mathematic

- Digital and Information manipulation
- Environment knowledge
- Social and Civic
- Artistic and cultural
- Learn to learn capacity
- Initiative and autonomy

5.1 Profile construction

First of all, we need to explain how will be designed and formed the subject and student profiles.

5.1.1 Subject profile construction

A subject needs that a student had developed certain set of competences if he or she wants to pass it. So, the subject profile will consist in the competences and de level of development that the subject needs to be studied correctly. For example, the subject Spanish Literature will need a high level of development for Linguistic and Arts competences and hardly in Mathematics competence.

So, we need to know which level of development each subject needs for each competence. This information was gathered by means of a questionnaire. We ask a big number of teachers from several High Schools in which degree (in a 1-5 scale) they think that subjects they teach need from students to have been developed each competence. Once we had questionnaire's results, next step was to aggregate them with a simple mean, in order to obtain a number which represent the level of development required for each competence. So, subject's profiles will remain as we can see in Table 4.

Competences /Subject	Mathematics	Spanish Literature	Social Sciences
Linguistic	1	5	3
Mathematic	5	1	2
Digital	3	2	2
Environment	2	2	5
Social	2	4	5
Arts	2	5	4
Learn to learn	4	3	3
Initiative	3	4	3

Table 4. Subject profiles constructed from competences

5.1.2 Student profile construction

In order to be consequent with subject profiles, student profiles must be built based on competences. In a specific moment each student has a degree of development for each competence. For example, it is known that some students outstand in Mathematics and Digital competences, but lack of Initiative and Arts ones, while on other students can happen just the opposite, or any other combination.

Spanish Academic System is about to evaluate by means of competences so that in a nearby future will be possible to know the level of development of each competence in any moment. However, in the present time this is not completely real. For this reason, in our proposal we

consider that student's competences should be weighted according to the mark obtained by student in the subjects has already studied. According to the discussion presented in section 3, where we said that subjects develop competences, in the same teacher's questionnaire presented before we ask them for specifying in which grade (in a 1-5 scale) each subject contribute to the development of a competence. Once we have got these results, we grouped subjects by grades in order to aggregate with a weighted average the percentage in which each subject contributes with each competence. Supposing that in a certain grade we have only three subjects, an example of the result of our quiz can be seen in Table 5.

Competences /Subject	Mathematics	Spanish Literature	Social Sciences
Linguistic	5%	70%	25%
Mathematic	80%	1%	19%
Digital	65%	15%	20%
Environment	25%	10%	65%
Social	5%	35%	60%
Artistic	10%	40%	50%
Learn to learn	30%	30%	40%
Initiative	20%	40%	40%

Table 5. Percentage of contribution of subjects in competences

From this table and students' marks in these subjects (for example, those on Table 6) we can calculate an approximation of the development level of development for each competence. The student profile can be seen on Table 7.

Subject/ Student	Mathematics	Spanish Literature	Social Sciences
John	10	1	3
Helen	2	10	8
Peter	4	5	10
Anne	9	6	9

Table 6. Example of students' marks in a 0-10 scale

Competences /Subject	John	Helen	Peter	Anne
Linguistic	0,98	4,55	3,10	3,45
Mathematic	4,29	1,61	2,58	4,49
Digital	3,63	2,20	2,68	4,28
Environment	2,28	3,35	4,00	4,35
Social	1,33	4,20	3,98	3,98
Artistic	1,45	4,10	3,70	3,90
Learn to learn	2,25	3,40	3,35	4,05
Initiative	1,80	3,80	3,40	3,90

Table 7. Student profiles

Now, with subjects and students profiles, we will see how a CBRS would work in order to make a recommendation.

5.2 Neighborhood formation

To provide recommendations, system needs to find those subjects that require a competence degree such that user had already developed. This is not exactly to find subjects with profiles equal to a student profile, but to find subjects profiles with required level of competences equal or lower that level in user profile. For example, looking at Table 4 we could think that a student with 5 in all competences will not be recommended any subject as he or she do not match well with the levels specified. However, the reality is that this student is able of studying any of them, because the requirements are fulfill enough.

To solve this question, we need a similarity measure which treat as well student which exceed requirements as students that simple fulfill them. For this reason we will not use similarity measure explained before, the cosine, and we will use a normalized variant of the Euclidean distance upgraded in order to follow this guideline, hence if student has a greater level for the desired competences, the similarity will give a positive result.

If student only need to fulfill the competence but system must take into account the rest of them, we will chose as competence development level the minimum between the student's value and de required for the subject so that it will be the same if student simply fulfill the requirement or overpass it anyway. This is achieved by using the minimum between the level required for the subject and the level accomplished by the student, instead of using only the level accomplished. This way the system treats equally those subjects with overpassed and simply fulfilled competence requirements, giving priority to those competence requirements not accomplished.

Consequently, let $r_{i,c}$ be the required level of development for competence *i* in subject *c*, and $v_{i,s}$ the computed value for student *s* in competence *i*. A is the amplitude of the value of development domain calculated as $A = b - a$, being *b* the top limit and *a* the bottom limit of such domain. In our case, interval of values used is between 1 and 5, both included.

$$u(c,s) = \frac{\sqrt{\sum_{i=1}^{n} \frac{(r_{i,c} - \min(r_{i,c}, v_{i,s}))}{A^2}}}{\sqrt{n}} \tag{3}$$

Equation 3 computes the similarity measure able of comparing student's capabilities and subject's requirements. The greater the value obtained the greater prepared will be the student to course that subject.

5.3 Top-N list generation

This equation will be applied to those subjects belonging to the target grade which student has to study next. The most often used technique for the generation of the top-*N* list is the one that select the *N* most valuated candidates with the similarity measure.

Once these *N* candidates are selected, they are presented as recommendation by the system.

6. Hybrid-OrieB: CB and CF in academic orientation

In section 4 was overviewed OrieB (Castellano 2009b), a web based Decision Support System built to support Spanish advisors in their task of helping students which modality to

choose in Baccalaureate, after finishing Secondary School. Specifically, the system aided advisors to obtain useful information about which subjects in each modality and which elective subjects suited better a student or which core subject might be extra hard. Thanks to this system advisors can develop their duties quicker and with reliable information.
However as it was pointed out, this system presented overall the new item problem which makes impossible to offer complete recommendations because in a continuous changing system, new subjects appears almost every year and CF is no able to support this kind of information. To solve this limitation and those seen in previous sections a new Hybrid-OrieB system has been built, using both CF and CB approaches in order to provide better recommendations and to overcome CF inherent limitations.

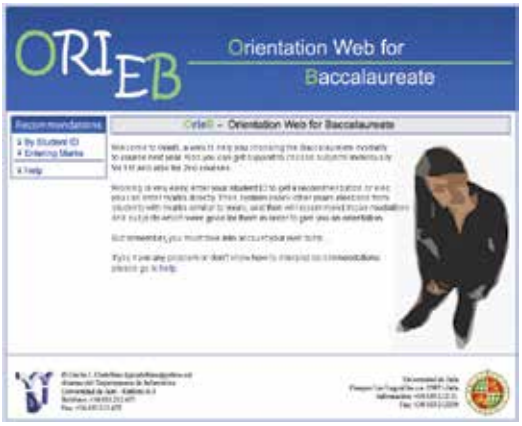


Fig. 3. Home Page of OrieB

Due to the importance that the information provided by this system can perform in the future decisions of students in early ages that they are not mature, we decided that it will be used just for advisors in order to support students but not directly by the students due to their lack of maturity.
The next step consists of choosing a hybridization method and presents the new system.

6.1 Hybridizing CF and CB for academic orientation

The main purpose of this chapter is to avoid the *new item* problem in OrieB by hybridizing CF and CB. When a new subject is presented in the system, it has no marks from students as it hasn't been studied yet by anybody. CF is unable to recommend this kind of subjects. This fact points out to the use of CB, as CB will always provide a recommendation for every target subject so that new subjects will not be a problem.
At this point, the system would obtain 2 lists of recommendations, one from CB and one from CF. Provided that CF uses 15 top subjects to elaborate its recommendation, we will set N for CB also in 15. So, we will have 2 lists of 15 subjects each with which we are going to work the hybridization.
In this sense, system will use subjects in both lists as follows:

- If dataset has no marks for a selected subject, CB recommendation will be used and the subject will be used to build the recommendation, because it is a *new subject*.
- If dataset presents any marks for that subject, system will weight recommendation of CF and CB using its own computed CF trust (Castellano 2009a, Castellano 2009b). Let

scf be the similarity computed by means of CF, *scb* the similarity computed by means of CB, and *t* trust computed for CF recommendation. New utility function for this subject will be computed as follows:

$$u'(c,s) = (scf * t) + \left(scb * \frac{1}{t}\right) \tag{4}$$

- If CF selects a subject not included by CB, it also will be used to perform the final recommendation because CF can offer serendipitous recommendations.

From previous procedure system will obtain a list with at the maximum 30 subjects in the case that CB recommends subjects that CF don't and vice versa. With this list system only has to order by similarity and recommend the same as done in the old OrieB. This way we have built a hybrid system by weighting and mixing results from separated CRS and CBRS.

6.2 Hybrid-OrieB

Once presented how system works internally in this section we will show what kind of recommendations and information OrieB offers and how to obtain it.

6.2.1 Interface

When advisors want to use OrieB to support a student, they just need to type the student ID (a unique number for each student which identifies him or her from others) or introduce the student's marks in the last year. The latter choice (Figure 4) offers the possibility of entering several marks instead of using all of them in order to obtain a more general orientation. But if not all marks are filled, the approximation of student's development level of competences can be not as accurate as desired. This way so the more marks filled the more accurate and customized will be the advices obtained by the system.

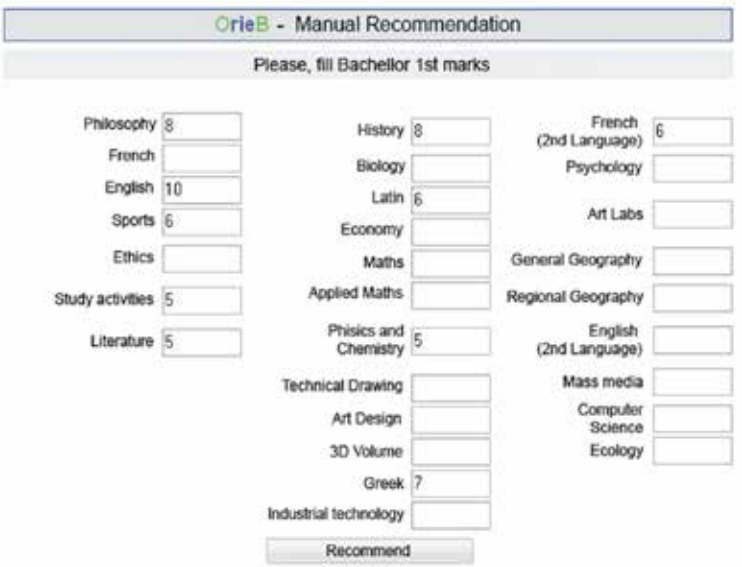


Fig. 4. Manual filling of marks

6.2.2 Supporting decisions

In order to help advisors in their tasks, OrieB offers three different types of support:

- Module recommendation
- Subject recommendation
- Warning difficulties in core subjects

Module or Vocational Program Recommendations

In order to aid advisors guiding students about the Module that better suits them according to their marks OrieB computes a Module recommendation based on a ordered list by interest (see Figure 5).

Each recommendation for a module incorporates an interest value and a trust value.

Interest value expresses the appropriateness of a module for the target student based on the predicted marks. System offers a linguistic term to make and explain recommendations than precise numerical values that can mislead the students in their decisions. So, OrieB will provide linguistic recommendations (Castellano 2009a).

Trust value shows the confidence about the previous interest value based on the ratio between the number of subjects whose predictions were obtained and the total number of subjects for the module, and the standard deviation of those predictions (Castellano 2009a).

Vocational Program Recommendation

Trust	Interest	Program
57%	Very High	Arts
60%	High	Humanities and Social sciences
64.22%	Medium	Natural sciences and health
54.5%	Very Low	Technology

Fig. 5. Vocational Program Recommendation

Support for choosing Elective and Module Subjects

Once students have chosen what module they prefer, they need to complete their curriculum with module and elective subjects. To support this decision OrieB offers separate recommendations for each group of subjects (see Figure 6).

Recommendation	Elective subject
Very high	Mass Media
High	Psichology
High	Computer science
Medium	French (2nd Language)

Fig. 6. Subject recommendation in OrieB

Warning Difficulties in Core Subjects

Finally, students also may need advices about what core subjects could be difficult for them. In this sense, the system offers a list with those core subjects with predictions lower than medium, it will warn the advisor which core subjects could cause difficulties to the student, with a trust level of the recommendation.

Core subject difficulty advising

Trust	Subject
74.33%	English
82%	Literature
16.67%	French

Fig. 7. Core subject difficulty advising

7. Concluding remarks

In this chapter we have introduced the problem of Academic Orientation, presented a system which make use of Collaborative Filtering techniques in order to recommend students an academic path by using their marks, and studying its limitations, proposed a hybrid CB and CF system in order to overcome the problems of *scarcity* and *new item* problem.

This system helps advisors in their task of supporting students and opens a matter of study in the Academic Orientation with CBE in which academic profiles are going to be more flexible and systems more capable of giving better recommendations for students in matter of improve and develop capabilities.

The origin of upgrading the support model for academic orientation with content based techniques is because of the recent adaptation of the Spanish Academic System to the Competence Based Education. This change provoked the appearance of new subjects and profiles that CF models in OrieB cannot managed because of the lack of data.

Consequently to overcome the *cold start* problem with these new subjects and due to the available information the more suitable model to achieve the goal of supporting advisors was the hybridizing with a content based model which provides a higher coverage but regarding the accuracy we need to wait to obtain real data sets. Additionally this upgraded version is ready to be extended and useful amid the ongoing changes of the academic system.

Eventually, we want to highlight though OrieB is performing pretty well, there exist different challenges that should guide our research in the future:

- OrieB does not take into account subjective information provided by students such as preferences, yet. So the system should be able to include not only information relative to their academic tour but also subjective and own information .
- Information provided by the system should not directly guide students because some reasoning about the results are necessary, so only advisors can use OrieB. More visual and self-explicative recommendations would be needed in this sense not only for allowing students using the system but also for providing advisors a better way of exploring and explaining academics alternatives to students.
- So far OrieB is focused on secondary and Baccaulerate grades. It seems interesting its extension to higher education.

8. References

Adomavicius, G. & Tuzhilin, A. (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering,* 17(6), 734-749.

Albadvi, A. & Shahbazi, M. (2009) A hybrid recommendation technique based on product category attributes. *Expert Systems with Applications,* 36(9), 11480-11488.

Breese, J. S., Hecherman, D. & Kadie, C. (1998) Empirical analysis of predictive algorithms for collaborative filtering, *Proceedings of Uncertainty in Artificial Intelligence. Fourteenth Conference.*, City.

Briñol, P., Falces, C. & Becerra, A. (2007) *ActitudesPsicolgía social.* (Madrid, McGraw Hill), 457-490.

Burke, R. (2000) *Knowledge-based recommender systems.* Encyclopedia of library and information systems. (vol. 69), 175-186.

Burke, R. (2002a) Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction,* 12(4), 331-370.

Burke, R. (2002b) *Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction.* (vol. 12)(Hingham, MA, USA, Kluwer Academic Publishers), 331-370.

Castellano, E. J. & Martínez, L. (2009a) *A linguistic recommender system for academic orientation*. IN B. S. R. (USA) (Ed.) Encyclopedia of E-Business Development and Management in the Global Economy.

Castellano, E. J. & Martínez, L. (2009b) A web-decision support system based on collaborative filtering for academic orientation. Case study of the spanish secondary school. *Journal of Universal Computer Science,* 15(14), 2786-2807.

Europe Tuning, E. S. (2000) *Una introducción a la contribución de las universidades al proceso de bolonia.*

Gardner, H. (Ed.) (1993) *Frames of mind: The theory of multiple intelligences.,* (New York, Basic Books).

Goldberg, D., Nichols, D., Oki, B. M. & Terry, D. (1992) Using collaborative filtering to weave an information tapestry. *ACM,* 35(12), 61-70.

González, J. & Wagenaar, R. (2008) *Universities' contribution to the bologna process, an introduction,* (2nd edn) (Bilbao, Deusto University).

Guttman, H. R. (1998) *Merchant differentiation through integrative negotiation in agent-mediated electronic commerce'*. Massachusetts Institute of Technology, School of Architecture and Planning.

Herlocker, J. L., Konstan, J. A., Borchers, A. & Riedl, J. (1999) An algorithmic framework for performing collaborative filtering, *Proceedings of 22nd International Conference on Research and Development in Information Retrieval*, City.

Herlocker, J. L., Konstan, J. A., Terveen, K. & Riedl, J. T. (2004) Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems,* 22(1), 5-53.

Horvath, T. (2009) A model of user preference learning for content-based recommender systems. *Computing and Informatics,* 28(4), 453-481.

Jung, K. Y., Park, D. H. & Lee, J. H. (2004) *Hybrid collaborative filtering and content-based filtering for improved recommender systemComputational science - iccs 2004, pt 1, proceedings.* (vol. 3036)(Berlin, Springer-Verlag Berlin), 295-302.

Krulwich, B. (1997) Lifestyle finder: Intelligent user profiling using large-scale demographic data. *AI Magazine,* 18(2), 37-45.

Lee, Y., Lee, S. & Lee, J. (2003) *Collaborative filtering recommendation system application based on stereotype model*, H. R. Arabnia & E. B. Kozerenko (Eds) *Mlmta'03: International*

conference on machine learning; models, technologies and applications. (Athens, C S R E a Press), 203-208.

Lenar, M. & Sobecki, J. (2007) Using recommendation to improve negotiations in agent-based systems. *Journal of Universal Computer Science,* 13(2), 267-286.

Martínez, L., Barranco, M. J., Pérez, L. G. & Espinilla, M. (2008) A knowledge based recommender system with multigranular linguistic information. *International Journal of Computational Intelligence Systems,* 1(3).

Martínez, L., Pérez, L. G. & Barranco, M. (2007a) A multi-granular linguistic content-based recommendation model. *International Journal of Intelligent Systems,* 22(5), 419-434.

Martínez, L., Pérez, L. G. & Barranco, M. (2007b) A multigranular linguistic content-based recommendation model:. *International Journal of Intelligent Systems,* 22(5), 419–434.

Pazzani, M. J. (1999) A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review,* 13(5-6), 393-408.

Perrenoud, P. (Ed.) (1999) *Building skills from school,* (Dolmen ed.).

Porcel, C., López-Herrera, A. G. & Herrera-Viedma, E. (2009) A recommender system for research resources based on fuzzy linguistic modeling. *Expert Systems with Applications,* 36(3), 5173-5183.

Resnick, P. & Varian, H. R. (1997) Recommender systems. *Communications of the ACM,* 40(3), 56-58.

Robertson, I. T. & Kinder, A. (1993) Personality and job competences: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology,* 66(3), 225-244.

Rodríguez, R. M., Espinilla, M., Sánchez, P. J. & Martínez, L. (2010) Using linguistic incomplete preference relations to cold start recommendations. *Internet Research,* 20(3), 296-315.

Romero, C., Ventura, S., Zafra A. & De Bra, P. (2009) Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems. *Computers and Education,* 53(3), 828-840.

Salgado, J. F. (1996) Personality and job competences: A comment on the robertson & kinder (1993) study. *Journal of Occupational and Organizational Psychology,* 69(4), 373-375.

Sarwar, B., Karypis, G., Konstan, J. & Reidl, J. (2001) Item-based collaborative filtering recommendation algorithms, *Proceedings of 10th international conference on World Wide Web*, City.

Schafer, J. B., Konstan, J. A. & Riedl, J. (2001) E-commerce recommender applications. *Data Mining and Knowledge Discovery,* 5(1-2), 115-153.

Sebastia, L., Garcia, I., Onaindia, E. & Guzman, C. (2009) E-tourism: A tourist recommendation and planning application. *International Journal on Artificial Intelligence* 18(5), 717-738.

Shevin, M. & Klein, N. K. (2004) The importance of choice-making skills for students with severe disabilities. *Research and Practice for Persons with Severe Disabilities,* 29(3), 161-168.

Symeonidis, P., Nanopoulos, A. & Manolopoulos, Y. (2007) Feature-weighted user model for recommender systems, *Proceedings of the 11th international conference on User Modeling*, City.

Takacs, G., Pilaszy, I., Nemeth, B. & Tikk, D. (2009) Scalable collaborative filtering approaches for large recommender systems. *Journal of machine learning research,* 10(623-656.

Zadeh, L. A. (1965) Fuzzy sets. *Information Control,* 8(338-353.

Zalba, E. a. G., N (2006) An approach to education based on competencies.

Zhen, L., Huang, G. Q. & Jiang, Z. H. (2010) An inner-enterprise knowledge recommender system *Expert Systems with Applications* 32(2), 1703-1712.

14

GMAA: A DSS Based on the Decision Analysis Methodology - Application Survey and Further Developments

Antonio Jiménez and Alfonso Mateos
Universidad Politécnica de Madrid
Spain

1. Introduction

Most of the real decision-making problems that we face nowadays are complex in the sense that they are usually plagued with uncertainty, and we have to take into account several conflicting objectives simultaneously. Decision support systems (DSS) play a key role in these situations helping decision makers (DMs) to structure and achieve a better understanding of the problem to make a final decision.

A number of multicriteria decision analysis techniques have underpinned the development and implementation of DSS in the last few decades (Figueira et al. (eds.), 2005). They include the *analytic hierarchy process* (Saaty, 1980; Pérez et al., 2006; Bana e Costa & Vansnick, 2008); *outranking methods*, especially *ELECTRE* (Roy, 1996; Figueira et al., 2005; Wang & Triantaphyllou, 2008); *PROMETHEE*, proposed in (Brans & Vincke, 1985; Brans et al., 1996; Brans & Mareschal, 2005; Mareschal et al., 2008), and their variants; or approaches based on *multi-attribute utility theory* (*MAUT*) (Keeney & Raiffa, 1993; Clemen 1996).

Corner and Kirkwood and Keefer et al. offer a systematic review of applications fields where DA methods were used and reported in operations research journals between 1970 and 2001 (Corner & Kirkwood, 1991; Keefer et al., 2004).

OR/MS Today published its tenth biennial decision analysis software survey in 2010 (Bucksaw, 2010). These surveys include broad-based information about available DSSs and their features.

The *decision analysis* (*DA*) methodology is widely used within *MAUT*. The goal of *DA* is to structure and simplify the task of making hard decisions as much as possible (Keeney & Raiffa, 1976; Clemen, 1996, Kirkwood, 1997). *DA* was developed on the assumption that the alternatives will appeal to the expert, depending on the expert preferences concerning the possible performances and the likelihood of each alternative performing in this manner.

What makes *DA* unique is the way in which these factors are quantified and formally incorporated into the problem analysis. Existing information, collected data, models and professional judgments are used to quantify the likelihoods of a range of performances, whereas utility theory is used to quantify preferences.

The usual or traditional approach to *DA* calls for single or precise values for the different model inputs, i.e., for the weight and utility assessments, as well as for the multi-attributed performances of the alternatives. However, most complex decision-making problems

involve imprecise information. Several reasons are given in the literature to justify why a DM may wish to provide incomplete information (Weber, 1987; Sarabando & Dias, 2010). For instance, regarding alternative performances, some parameters of the model may be intangible or non-monetary as they reflect social or environmental impacts. Also, the performances may be taken from statistics or measurements, and the information that would set the value of some parameters may be incomplete, not credible, contradictory or controversial. At the same time, it is often not easy to elicit precise weights, which may be represented by intervals, probability distributions and even fuzzy values, or just satisfying ordinal relations. DMs may find it difficult to compare criteria or not want to reveal their preferences in public. Furthermore, the decision may be taken within a group, where the imprecision of the preferences is the result of a negotiation process.

This situation where it is not possible to indicate precise values for the parameters and quantities involved is often referred to as *decision-making with imprecise information*, *with incomplete information* or *with partial information*, together with *incomplete knowledge* or *linear partial information* (Kmietowicz & Pearman, 1984; Kirkwood & Sarin, 1985; Hazen, 1986; Ríos Insua & French, 1991).

A lot of work on *MAUT* has dealt with incomplete information. (Sage & White, 1984) proposed the model of imprecisely specified *ISMAUT*, where preference information about both weights and utilities is assumed not to be precise. (Malakooti, 2000) suggested a new efficient algorithm for ranking alternatives when there is incomplete information about the preferences and the performance of the alternatives. This involves solving a single mathematical programming problem many times. (Ahn, 2003) extended Malakooti's work. (Eum et al., 2001) provided linear programming characterizations of dominance and potential optimality for decision alternatives when information about performances and/or weights is incomplete, extended the approach to hierarchical structures (Lee et al., 2002; Park, 2004), and developed the concepts of *weak potential optimality* and *strong potential optimality* (Park, 2004). More recently, (Mateos et al., 2007) considered the more general case where imprecision, described by means of fixed bounds, appears in alternative performances, as well as in weights and utilities.

The s*tochastic multicriteria acceptability analysis* (*SMAA*) and *SMAA-2* methods (Lahdelma & Salminen, 1998; 2001) were developed for support in discrete group decision-making problems, where weight information is absent. These methods explore the weight space in order to describe the ratings that would make each alternative the preferred one. This situation was also considered by other authors (Bana e Costa, 1986; Charnetski & Soland, 1978; Nijkamp et al., 1990; and Voogd, 1983).

(Sarabando & Dias, 2010) gives a brief overview of approaches proposed by different authors within the *MAUT* and *MAVT* (multi-attribute value theory) framework to deal with incomplete information.

2. The GMAA decision support system

The *generic multi-attribute analysis* (GMAA) *system* is a PC-based DSS based on an additive multi-attribute utility model that is intended to allay many of the operational difficulties involved in the DA cycle (Jiménez et al., 2003; 2006).

This cycle can be divided into four steps: 1) structuring the problem (which includes specifying objectives, building a value hierarchy and establishing attributes for the lowest-level objectives); 2) identifying the feasible alternatives, their performances and uncertainty

(if necessary); 3) quantifying preferences (which includes the assessment of the component attribute utilities, as well as the weights representing the relative importance of criteria); and 4) evaluating alternatives and performing sensitivity analysis (SA).

The GMAA system accounts for uncertainty about the alternative performances. Quantifying preferences involves assessing component utilities, which represent the DM's preferences over the possible attribute performances, and eliciting weights, which account for the relative importance of criteria.

The GMAA system provides four procedures for assessing component utilities: 1) construct a piecewise linear imprecise utility function, 2) construct an imprecise utility function with a gamble-based method (Jiménez et al., 2003), 3) assign imprecise utilities to discrete attribute values and 4) directly provide subjective values.

There are two main ways of representing the relative importance of criteria. The first is to preemptive order attributes (Adelbratt & Montgomery, 1980), and the second is to use attribute weights. The second option is more widespread and is used in the GMAA system. Different methods have been proposed to elicit weights by different authors, such as *DIRECT point allocation*, *simple multi-attribute rating technique - SMART -* (Edwards, 1977), *SWING weighting* (von Winterfeldt & Edwards, 1986; Edwards & Barron, 1994), *SMART using swings - SMARTS -* (Edwards & Barron, 1994), *SMART exploiting ranks - SMARTER* – Edwards & Barron, 1994), *TRADE-OFFS weighting* (Keeney & Raiffa, 1993), *pricing out method* (Keeney & Raiffa, 1993), *analytic hierarchy process* -AHP- (Saaty, 1980), or *preference programming* (Salo & Hämäläinen, 1995). The GMAA system provides *DIRECT point allocation* and *TRADE-OFFS weighting*.

The GMAA system accounts for incomplete information about the DM's preferences through value intervals as responses to the probability questions that the DM is asked, leading to classes of utility functions and weight intervals. This is less demanding for a single DM and also makes the system suitable for group decision support, where individual conflicting views in a group of DMs can be captured through imprecise answers.

An additive multi-attribute utility function is used to evaluate the alternatives, taking the form

$$u(S^j) = \sum_{i=1}^{n} w_i u_i(x_i^j) \tag{1}$$

where w_i is the ith attribute weight, x_i is the performance for alternative S^j in the ith attribute and $u_i(x_{ij})$ is the utility associated with the above performance.

For the reasons described in (Raiffa, 1982; Sterwart, 1996), the additive model is considered to be a valid approach in most practical situations. It is used to assess, on the one hand, average overall utilities, on which the ranking of alternatives is based and, on the other, minimum and maximum overall utilities, which give further insight into the robustness of this ranking.

The GMAA provides several types of SA. It can assess the *stability weight interval* for any objective at any level in the hierarchy. This represents the interval where the average normalized weight for the considered objective can vary without affecting the overall ranking of alternatives or just the best ranked alternative.

On the other hand, the *assessment of non-dominated and potentially optimal alternatives* (Mateos et al., 2007) takes advantage of the imprecise information gathered during the assignment of

component utilities and weights and the performance of the entered uncertain alternatives to definitely reject poor alternatives, mainly by discarding dominated and/or non-potentially optimal alternatives.

The GMAA system computes the potentially optimal alternatives among the non-dominated alternatives because these are alternatives that are best ranked for at least one combination of the imprecise parameters, i.e., weights, component utility functions and alternative performances.

Finally, *Monte Carlo simulation techniques* enable simultaneous changes of the weights and generate results that can be easily analyzed statistically to provide more insight into the multi-attribute model recommendations (Mateos el at, 2006). While the simulation is running, the system computes several statistics about the rankings of each alternative, like mode, minimum, maximum, mean, standard deviation and the 25th, 50th and 75th percentiles. This information can be useful for discarding some available alternatives, aided by a display that presents a multiple boxplot for the alternatives.

The GMAA system provides three general classes of simulation. In the *random weights* option, weights for the attributes are generated completely at random, which means that there is no knowledge whatsoever of the relative importance of the attributes. In the *rank order weights* option, attribute weights are randomly generated preserving a total or partial attribute rank order, which places substantial restrictions on the domain of possible weights that are consistent with the DM's judgement of criteria importance, leading to more meaningful results. Finally, in the *response distribution weights* option, attribute weights are now randomly assigned values taking into account the normalized attribute weight intervals provided by the DM in the weight elicitation methods.

The Universidad Politécnica de Madrid registered the GMAA system and a free version (installation package and user's guide) is available for academic purposes at http://www.dia.fi.upm.es/~ajimenez/GMAA.

3. Real application to complex decision-making problems

GMAA has proved to be useful for solving complex decision-making problems in different areas. We summarize below some of the problems in which GMAA was used as a key part of the decision-making process.

3.1 Selection of intervention strategies for the restoration of aquatic ecosystems contaminated by radionuclides

The first problem in which the GMMA system was used was to evaluate intervention strategies for the restoration of aquatic ecosystems contaminated by radionuclides.

This problem was studied in depth as part of several European projects in which we participated: MOIRA (A model-based computerized system for management support to identify optimal remedial strategies for restoring radionuclide contaminated aquatic ecosystem and drainage areas) (Monte et al., 2000), COMETES (implementing computerized methodologies to evaluate the effectiveness of countermeasures for restoring radionuclide contaminated freshwater ecosystems) (Monte et al., 2002), EVANET-HYDRA (evaluation and network of EC-decision support systems in the field of hydrological dispersion models and of aquatic radioecological research), and EURANOS (European approach to nuclear and radiological emergency management and rehabilitation strategies).

Throughout these projects, a synthetic, flexible and user-friendly computerized decision support system, MOIRA, was implemented and tested on several real scenarios. The system included a multi-attribute analyses module for the global assessment of the effectiveness of the intervention strategies. This module was the origin of the GMAA system, which was finally built into the last versions of the MOIRA system.

The selection of intervention strategies was based on environmental models for predicting the migration of radionuclides through freshwater and coastal ecosystems and the effects of feasible countermeasures on contamination levels. Moreover, other social and economic criteria were taken into account.

Several real scenarios contaminated as a consequence of the Chernobyl accident were analysed, like lake Øvre Heimdalsvatn (Jiménez et al., 2003), located in Oppland county (Norway), lake Kozhanovskoe (Ríos Insua et al., 2004), located in the region of Bryansk (Russia), and lake Svyatoye in Belarus (Ríos Insua et al. 2006).

Lake Svyatoye is located 237km from the Chernobyl nuclear power plant and 30km southeast of Kostyukovichy (Belarus), and the ^{137}Cs contamination in the area was over 1480 kBq/m^2. It is a collapse (karst) type lake. The maximum depth, at the center of the lake, is 5.2 m. It has both precipitation and subsurface water supply and the water balance is controlled by evaporation and subsurface runoff.

According to the DA cycle, we started to build an objective hierarchy including all the key aspects to be considered in the problem, see Fig. 1.

Environmental Impact (Environ. Imp) is one of the main objectives of the decision analysis. It was divided into *Lake Ecosystem Index* (L.E.I.), a simple and rational approach for measuring the ecological status of a lake, and *Radiation Dose to Biota* (Dose to Fish). *Social Impact* (Social Imp.) was handled by two sub-objectives: minimizing impact on health (Dose to Man) and *Living Restrictions* (Living restr).

Regarding dose to man, we focused on the *Dose to Critical Individuals* (Dose Crit In), who should never receive radiation levels above thresholds for early health effects, and *Collective Dose* (Coll. Dose), which was linearly related to the increase in the risk of developing serious latent effects, mainly cancers. As regards living restrictions, other impacts were taken into consideration. These include countermeasures affecting the direct consumption of fish for food or its processing in the food industry, drinking water and water used by the food industry, the use of water for crops irrigation and the recreational uses of water bodies. For all these objectives, the attributes were the amount of fish affected by restrictions (Amount fish), as well as the duration of such restrictions (Ban Duration).

Finally, *Economic Impact* (Economic Im) was divided into *Direct Effects* (Direct Eff.), more amenable to quantification, and *Intangible Effects* (Intang. Eff.), like loss-of-image and adverse market reactions for the concerned area, which could also be subjectively valued by the user. The direct effects include the costs generated by the different bans or restrictions to normal living conditions, which can be sub-divided into *Costs to the Economy* (Cost to econ) and the more subjective costs of lost recreation, and *Application Costs* (Applic. Cost), i.e., costs of chemical and physical remedial countermeasures.

Taking into account expert knowledge and opinions, a set of seven intervention strategies were proposed for analysis, combining chemical countermeasures with fishing bans so as to reduce the radiological and environmental impact:

- *No action* (natural evolution of the situation without intervention).
- *Potassium* (15 tonnes of potassium chloride added to the lake in April 1987).

- *Fertilizer* (800 kg of fertilizer added to the lake between April and July 1987-90).
- *Food ban* (fish consumption ban when ^{137}Cs content in fish is >1000 Bq/kg).
- *Lake liming* (15 tonnes of lime added to the lake in April 1987).
- *Sediment removal* (250,000m^2 of sediments down to a depth of 10cm removed from the lake, between May–June 1990, i.e., 125,000 m^2/month).
- *Wetland liming* (30 tonnes of lime added to the catchment in May 1987).

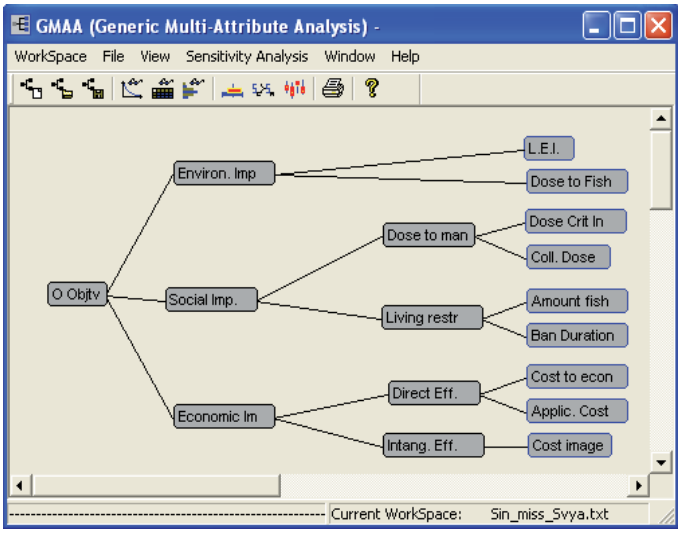


Fig. 1. Objectives hierarchy for lake Svyatoye

The impacts/performances of the intervention strategies were then established in terms of the attributes associated with the lowest-level objectives and described under uncertainty by vectors of values (see Table 3 in Ríos- Insua et al., 2006).

DM preferences were elicited according to *DA* cycle. An imprecise component utility function was assessed for each attribute, representing DM preferences concerning the respective possible attribute impacts. Fig. 2 shows the class of utility functions for the *Dose to Critical Individuals*.

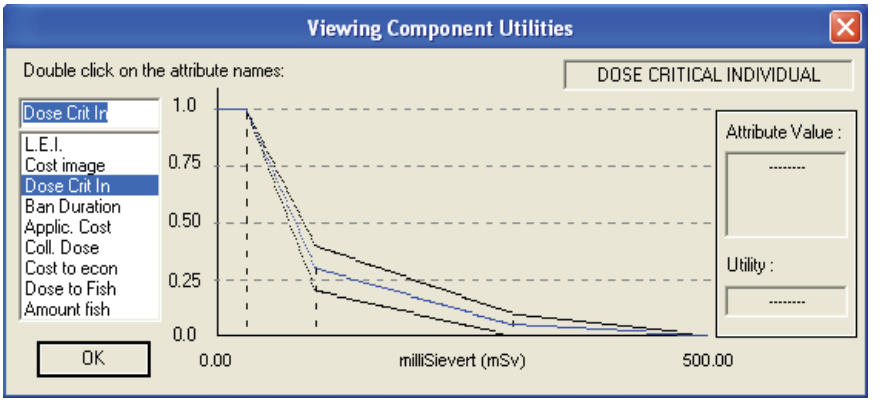


Fig. 2. Classes of utility functions for *Dose to Critical Individuals*

On the other hand, objective weights representing their relative importance were elicited along the branches of the objectives hierarchy. Then, the attribute weights used in the additive multi-attribute utility model were assessed by multiplying the elicited weights in the path from the overall objective to the respective attributes, see Fig. 3. These attribute weights are indicators of the influence of the individual criteria on the decision.
The additive multi-attribute utility model, which demands precise values, was then used to assess, on the one hand, average overall utilities, on which the ranking of alternatives is based and, on the other, minimum and maximum overall utilities, which give further insight into the robustness of this ranking.

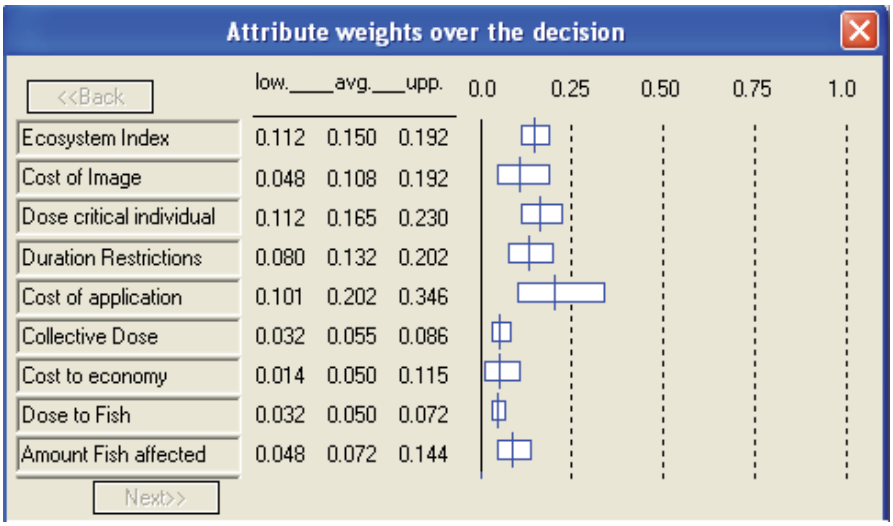


Fig. 3. Imprecise attribute weights in the selection of intervention strategies

Fig. 4 shows the ranking of the intervention strategies for lake Svyatoye, where the vertical white lines on each bar represent average utilities. The best-ranked intervention strategy was *Potash* with an average overall utility of 0.802, followed by *Lake Liming* (0.754) and *Wetland Liming* (0.751), whereas the worst ranked option was *Sediment Removal* with a utility of 0.607.

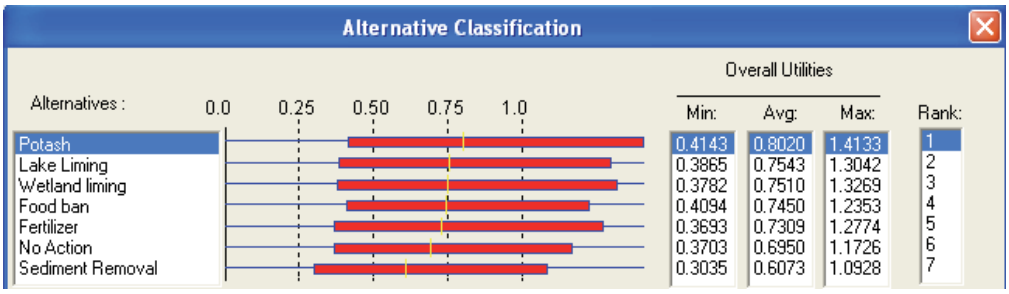


Fig. 4. Ranking of intervention strategies for lake Svyatoye

Looking at the overlapped utility intervals (robustness of the ranking of strategies), however, we concluded that the information obtained by this evaluation was not meaningful enough to definitively recommend an intervention strategy. Consequently,

sensitivity analysis should be carried out to output further insight into the recommendations.

First, all the intervention strategies were non-dominated and potentially optimal, i.e., they were not dominated by any other strategy and best-ranked for at least one combination of the imprecise parameters, i.e., weights, component utilities and strategy impacts. Thus, we could not discard any of them from further consideration.

Then, Monte Carlo simulation techniques were applied. The *response distribution weights* option was selected, i.e., attribute weights were randomly assigned values taking into account the weight intervals provided by the DMs in weight elicitation, see Fig. 3. Fig. 5 shows the resulting multiple boxplot.

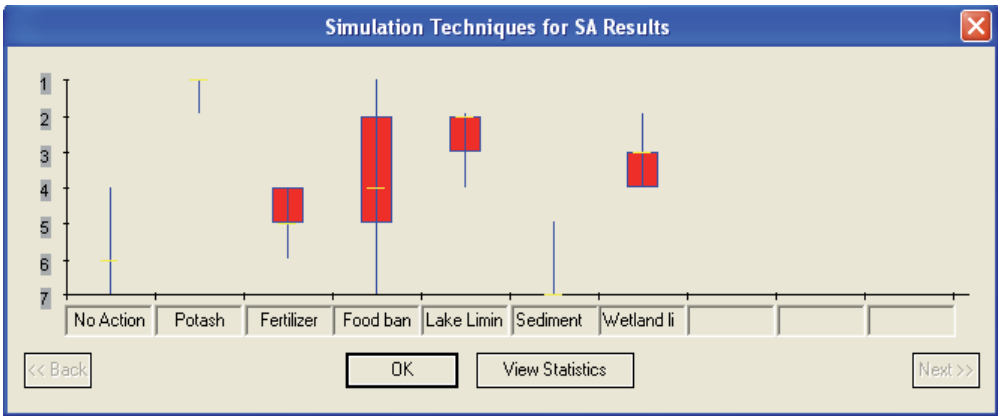


Fig. 5. Multiple boxplot with response distribution weights

Only two intervention strategies were best ranked for at least one combination of weights, *Potash* and *Food Ban*, but the worst classification for both was second and seventh, respectively. Moreover, the mean classifications were 1.065 and 3.785, respectively, indicating that the intervention strategy that should be definitively recommended was *Potash*.

3.2 Selection of a supplier for cleaning services in an underground transportation company

The GMAA system was also used to select a supplier for cleaning services in a European public underground transportation company (Jiménez et al., 2007).

From a cost management perspective, the cleaning service represents a sizeable part of the operating costs of an underground transportation service, accounting for close to 10% of total annual operating expenses. Inherent service intangibility, heterogeneity and inseparability greatly constrain service acquisition and management (Zeithaml et al., 1990). For the purpose of determining the most economically advantageous suppliers, several conflicting criteria were taken into account simultaneously to provide the most relevant information about what other factors, in conjunction with prices, to apply during the award-of-contract process, like delivery conditions and human resources, technical merit and resources, quality control procedures, etc., according to the public procurement policies and legislation for national and European public organizations. These criteria were compiled as a result of a team effort, including procurement and technical experts from the

organizational areas responsible for cleaning services, finance and other departments or representatives, such as legal affairs and customer service.

An objective hierarchy with the following five main top-level objectives was built: *Delivery conditions and human resources*, which accounts for how consistent and coherent the human resources allocated to the services are; *Technical merit and resources*, which it is an important efficiency factor leading to a significant reduction in labor cost; *Price*, which represents the lowest price offered by suppliers; *Quality control procedures*, which accounts for accredited quality certifications and how quality systems and procedures are deployed; and *Graffiti prevention and cleanup*, which is one of the most common incidents detracting from the appearance of the underground buildings. Twenty-one lowest-level objectives were included in the above objective hierarchy, see Fig. 6, and the corresponding attributes were established to indicate to what extent they were achieved by the respective offers.

Following the European Community directives, stating that criteria should be set and at least ordered by importance before examining the offers, DM preferences were quantified before identifying offers. Fig. 7 illustrates attribute weights. The most important attribute in the decision was *Price*, with an average normalized weight of 0.513, followed at a distance by workload, quantity of technical means, number of quality certifications, experience with graffiti and workload allocation, with average normalized weights of 0.125, 0.082, 0.052, 0.052 and 0.036, respectively.

Next, the feasible supplier offers were identified and measured in terms of the 21 attributes. The performances for the six offers considered and the component utilities assessed are reported in (Jiménez et al., 2007). Note that uncertainty about some of the performances was accounted for by means of percentage deviations, which represent tolerances or correction factors for offers where information provided by the bidders was somewhat ambiguous or was not clear enough.

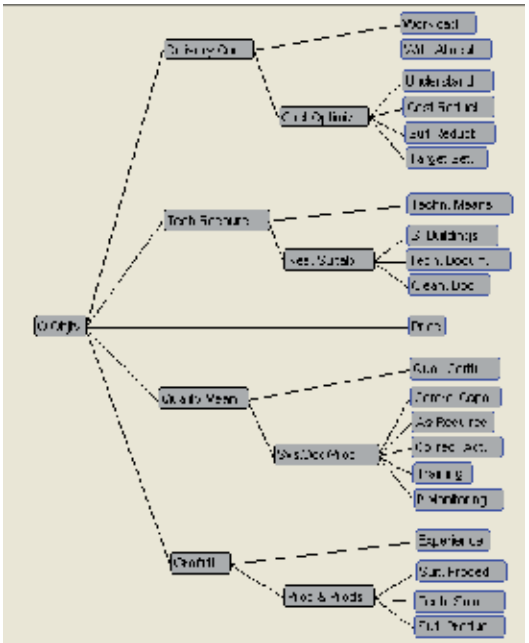


Fig. 6. Objective hierarchy in the selection of a supplier for cleaning services

The additive model was then used to evaluate the offers under consideration, see Fig. 8. Offers 3 and 6 were the best and worst ranked tenders. Looking at the utility intervals, we could discard offer 6, because its maximum utility was lower than the minimum utility of offer 2. Consequently, we concluded that it was dominated. Although offer 3 appeared to be the most highly recommended, however, the overlapped utility intervals were examined in more detail through SA.

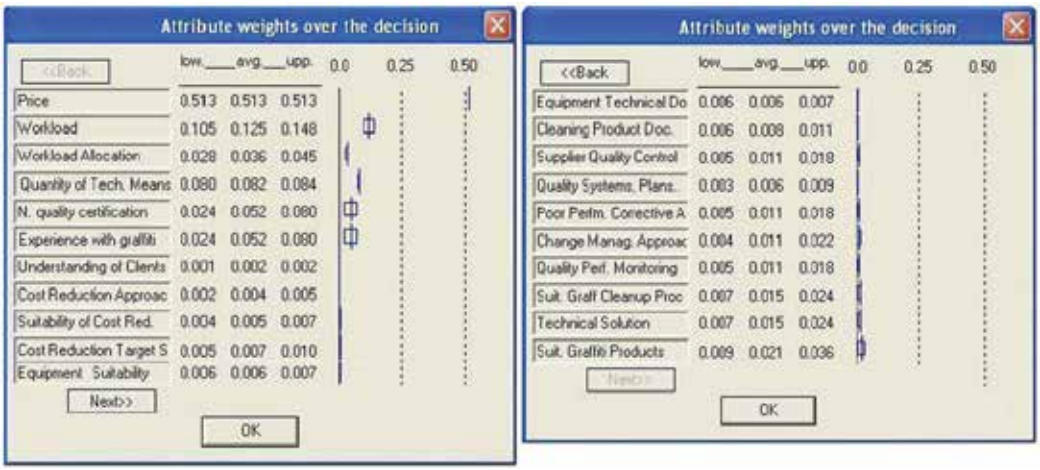


Fig. 7. The attribute weights in the selection of a supplier for cleaning services

First, the *stability interval* for all objectives throughout the hierarchy were computed, i.e., the interval where the average normalized weight for any objective can vary without the best ranked offer changing. The resulting stability intervals for all objectives throughout the hierarchy were [0, 1], which meant that, whatever their relative importance, offer 3 remained the best-ranked tender, except for *Delivery conditions and human resources*, and *Quality control procedures* with stability weight intervals [0, 0.598] and [0, 0.748], respectively. Taking into account that the narrower a stability weight interval is the more sensitive the offers ranking is, we concluded that the offers ranking was robust regarding the elicited weights.

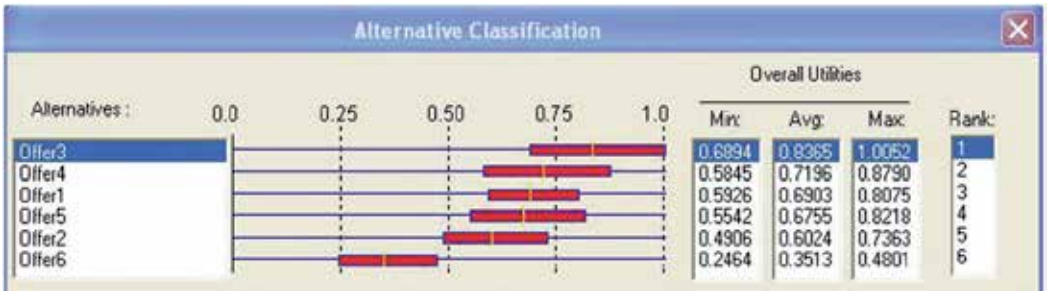


Fig. 8. Offer evaluation

On the other hand, only offer 3 was non-dominated and potentially optimal, so we could definitively conclude that it is the best offer. Monte Carlo simulation techniques were performed led to the same finding; offer 3 was the best.

3.3 Selection of segment to be sourced from low cost countries for a global industrial equipment manufacturer

Another complex decision-making problem in which the GMAA system was to select segments to be sourced from low cost countries for a global industrial equipment manufacturer (Jiménez et al., 2005).

Competitive pressures are forcing companies to reduce their overall costs, while delivering faster and more diverse product portfolios to be more responsive to customers and competitors. In response to these pressures, companies are increasingly taking advantage of the opportunity to source from low cost countries (LCC) to achieve significant savings and give their organizations a competitive advantage. For a global industrial equipment manufacturer with material costs accounting for about 50% of the value of its final products, sourcing performance is crucial to original equipment manufactured (OEM) competitiveness.

Even though multinational companies have been sourcing from LCC for many years, purchasing in these regions is often very risky, and many companies spend a lot of time and energy trying to identify and minimize these risks (identifying reliable sources, political instability, currency risks, longer lead-times, more complex logistics, different/non-existent legal structures…).

Typically, incremental cost reductions of 15%-20% can be achieved by sourcing from LCC. Before moving the supply source for some specific segment categories to these regions, however, the segments have to be proven to have a comprehensive risk assessment, balanced against potential for lower costs. Although benefits are compelling, they come with significant challenges.

For the purpose of determining segment categories with the highest profit potential for sourcing from LCC, a range of conflicting criteria were taken into account simultaneously. Therefore, the promise of significant cost reductions was not the only consideration, and the country, industry and supplier risks were key factors considered during the prioritization of the category segments. In this case, the responsible organization of procurement evolved into a formal decision process, and other strategic issues related to LCC sourcing activities were quantified and formally incorporated into the analysis. This way, cost-cutting potential was only one of the purchaser's objectives.

As shown in Fig. 9, the Overall Objective (O.Objtv) was split into two main sub-objectives: *Potential Benefits* (Pot. Benefit) and *Risks* (Risk). *Potential Benefits* were measured in terms of four sub-objectives. The *Total annual expenditure* (Spend HCC) on all parts in the segment not sourced from LCC. The expenditure is an indicator of the potential volume with which we are dealing. The higher the expenditure is, the more room there is for savings. The *Price per kg* (Price Kg.) indicates the price regarding the value-added for the parts produced in high cost countries (HCC). A higher HCC price/kg value-added represents a high potential benefit. The *Factor cost content* (F C Content) is subject to comparison between HCC and LCC. Labor is the main factor cost to be taken into account. The higher the labor content is, the larger the cost difference window between sourcing countries is. High labor content represents potential high cost savings when sourcing from LCC. Finally, *Supplier switching costs* (Sup. S Costs) is the cost of switching from the current supplier set-up to a new supplier. The higher the switching cost, the lower the potential benefit. Tooling cost is the most important and most easily quantifiable switching cost to be accounted for. Other switching costs can be considered if known.

On the other hand, *Risks* is split into four sub-objectives. *Complexity of parts* (Complx Parts) represents part of the risk of selecting a new supplier. Technical issues related to quality and material specification could be added to the assessment of the total complexity of parts in each segment. A higher complexity implies a higher risk. *Risk with current suppliers* (Risk C Suppl) quantifies the number of segments that the supplier is supplying at that moment. Moving business in one segment from the current supplier to LCC will influence the supply of the other segments (price increases, production stop pages, low performance, etc.). Therefore, the more segments supplied by one supplier, the higher the risk of moving to LCC.

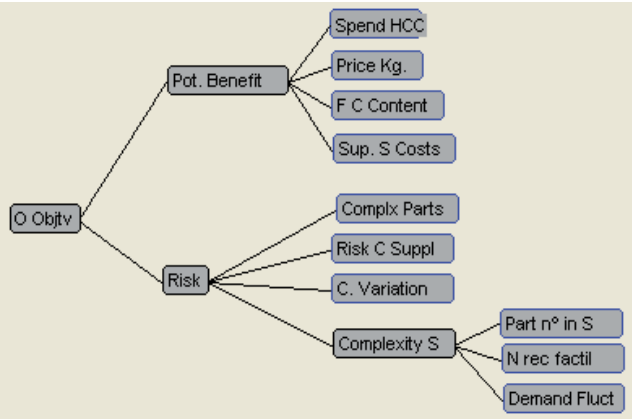


Fig. 9. Objective hierarchy to create a cost efficient production process

The *Coefficient of variation* (C. Variation) tells us how homogeneous the price per kg of the parts in the segment is. The higher the coefficient of variation, the greater the risk, because the handling of the different parts of the segment varies more. Finally, *Complexity of segments* (Complexity S) represents supply chain issues in relation to the purchase of parts from a wider perspective. The *Number of parts within a segment* (Part nº in S), the *Number of receiving facilities* for the parts in the segment (N rec facil) and *Demand fluctuation* (Demand Fluct) are the main quantifiable criteria to be taken into consideration. A set of 19 non-metallic product segments was identified. Their performances in terms of the 10 attributes under consideration are reported in (Table 2, Jiménez et al., 2005).

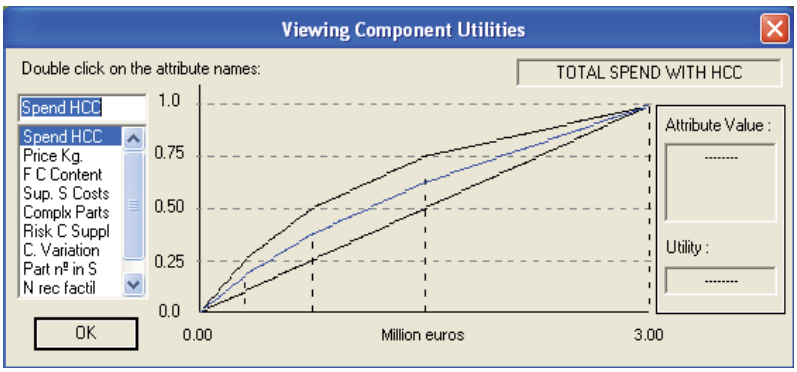


Fig. 10. Imprecise linear piecewise utility function for the *Total annual expenditure*

Fig. 10 shows the assessed imprecise linear piecewise utility function for the *Total annual expenditure* (Spend HCC), whereas Fig. 11 shows the attribute weights for the decision.
Fig. 12 shows the ranking of segments derived from the additive multi-attribute utility model. *SG19*, *SG11*, *SG18* and SG*13* are the best-ranked segments, with average overall utilities of 0.6963, 0.6835, 0.5877 and 0.5417, respectively; whereas *SG9*, *SG4* and *SG5* are the worst ranked segments, with average overall utilities of 0.3833, 0.3716 and 0.3213. Although *SG19* appears to be the most highly recommended segment, sensitivity analysis (SA) should be used to examine the overlapped utility intervals (ranking robustness) in a more detail.

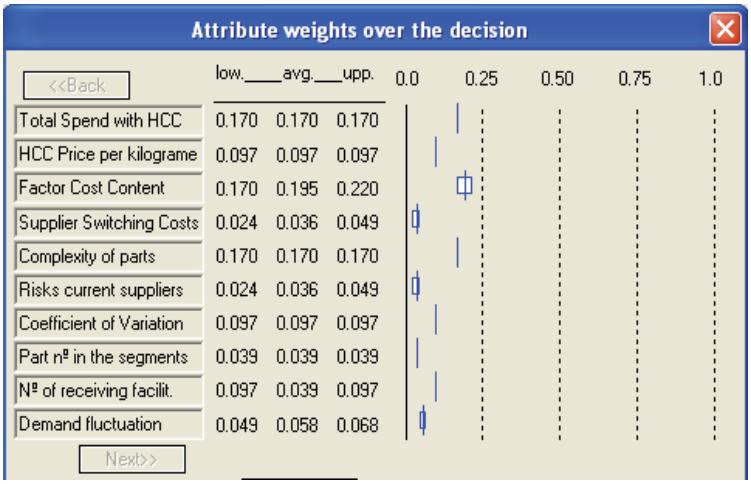


Fig. 11. Attribute weights in the selection of segment to be sourced from low cost countries

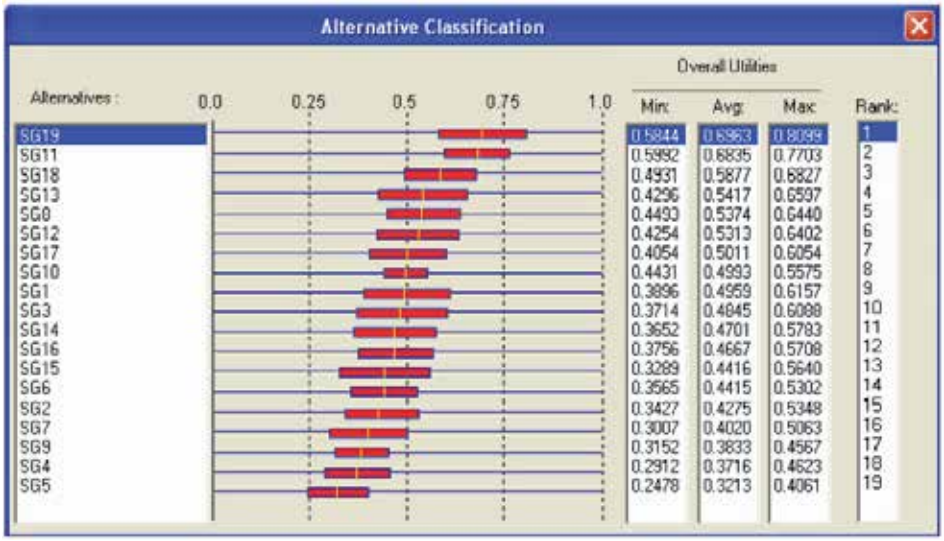


Fig. 12. Ranking of segment to be sourced from low cost countries

Only three segments, *SG11*, *SG18* and *SG19*, are non-dominated and potentially optimal. Note that these were the best-ranked segments. Looking at the box plots output by Monte

Carlo simulation techniques for *SG11*, *SG18* and *SG19*, we found that they are always ranked second, third and first, respectively, see Fig. 13. Therefore, we concluded that the segment category with the best tradeoff between potential benefit and risks to be sourced from LCC was *SG19: Hand lay-up composite parts*. However, we were not just interested in the best segment to be sourced from LCC, our aim was to identify a segment set with a good enough tradeoff between potential benefit and risk.

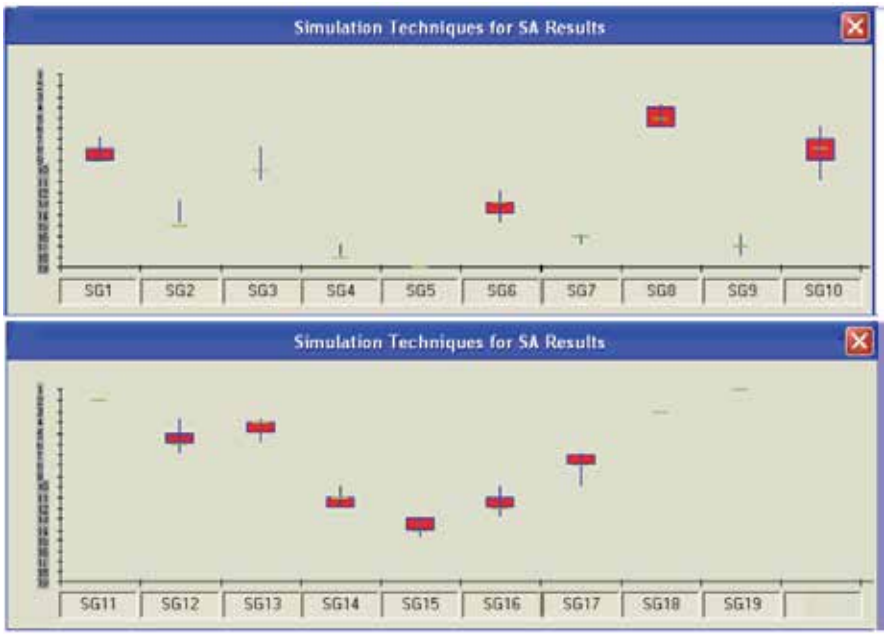


Fig 13. Results of Monte Carlo simulation techniques for the considered segments

Rank	Alternatives	Overall Utilities Min:	Avg:	Max:	Total Spend w/HCC	% spend	Cumm %
1	SG19	0.584421	0.696259	0.809945	2,30	6,78%	6,78%
2	SG11	0.599205	0.683484	0.770322	5,10	15,03%	21,81%
3	SG18	0.493066	0.587714	0.682714	3,94	11,61%	33,42%
4	SG13	0.429618	0.541691	0.659655	0,95	2,80%	36,22%
5	SG08	0.449297	0.537422	0.643996	0,75	2,21%	38,43%
6	SG12	0.425394	0.531276	0.640246	1,50	4,42%	42,85%
7	SG17	0.405360	0.501112	0.605404	2,49	7,34%	50,19%
8	SG10	0.443123	0.499324	0.557503	4,75	14,00%	64,19%
9	SG01	0.389639	0.495854	0.615723	1,73	5,10%	69,29%
10	SG03	0.371448	0.484540	0.608774	0,72	2,12%	71,41%
11	SG14	0.365216	0.470133	0.578293	0,64	1,89%	73,30%
12	SG16	0.375603	0.466701	0.570811	1,58	4,66%	77,95%
13	SG15	0.328941	0.441557	0.564017	1,49	4,39%	82,35%
14	SG06	0.356492	0.441473	0.530208	0,91	2,68%	85,03%
15	SG02	0.342685	0.427510	0.534785	1,07	3,15%	88,18%
16	SG07	0.300710	0.401960	0.506272	0,63	1,86%	90,04%
17	SG09	0.315248	0.383333	0.456682	0,65	1,92%	91,95%
18	SG04	0.291185	0.371589	0.462338	1,20	3,54%	95,49%
19	SG05	0.247849	0.321267	0.406146	1,53	4,51%	100,00%

Fig. 14. Final recommendation

Taking into account the above segments' rankings and the results of SA, the OEM management finally recommended the best ranked segments, accounting for the 60% of the total expenditure on non-metallic category segments, to be sourced from LCC, see Fig. 14.

3.4 Selection of intervention strategies against eutrophication and the drastic decrease in the bird population in a fjord

DSS are also becoming increasingly popular in environmental management (Stam et al., 1998; Tecle et al., 1998; Ito et al., 2001; Poch et al. 2004). Examples where MAUT has been used for environmental management problems can be found, e.g., in the field of forest management (Ananda & Herath, 2009), natural resource management (Mendoza & Martins, 2006)), different fields of water management (Linkov et al., 2006; Joubert et al., 2003), river management (Reichert et al., 2007; Corsair et al., 2009), landscape ecology (Geneletti, 2005), evaluation of farming systems (Prato & Herath, 2007), and site selection for hazardous waste management (Merkhofer et al., 1997).

Many coastal zones all over the world have been put under increasing pressure from human activities over recent decades, through overfishing, nutrient inputs, and global warming. A related challenge is to come up with models that can accurately predict ecosystem changes from human activity and preventive or remedial actions. In this sense, the GMAA was also used to select intervention strategies against eutrophication and the drastic decrease in the bird population in Ringkøbing Fjord (Bryhn et al., 2009; Jiménez et al., 2011).

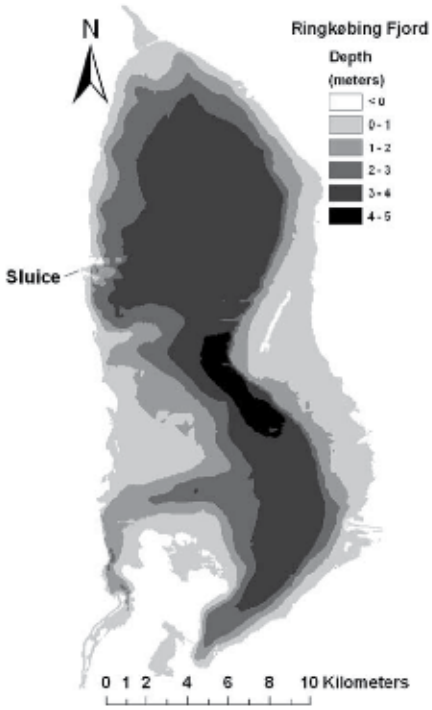


Fig. 15. Ringkøbing Fjord

Ringkøbing Fjord is a large and shallow brackish lagoon on the west coast of Denmark. It has an area of 300 km^2, a volume of 0.57 km^3, a maximum depth of 5.1 m and a mean depth

of 1.9 m. The lagoon receives large (2 km^3yr^{-1}) freshwater inputs from the catchment, as well as saltwater inputs through a sluice that connects the lagoon to the sea, see Fig. 15.

Ringkøbing Fjord has gone through two environmental regime shifts during the last decades (Håkanson et al., 2007), which has stirred up public sentiment in the area, mainly because of the disappearance of waterfowl.

The nine considered intervention strategies consisted of the abatement of nitrogen and/or phosphorous, the construction and maintenance of a saltwater dumping station or a second sluice to increase the water exchange between the lagoon and the outside sea and some combinations of the above. All of them were aimed at achieving good water quality in terms of trophic state and conditions for waterfowl and were compared with the no action alternative and with decreasing the salinity level to 7.2 per thousand, the mean value in the lagoon during the 12-year period preceding a major structural regime shift in 1996 (Håkanson & Bryhn, 2008; Petersen et al., 2008).

Intervention strategies were evaluated considering their environmental, social and economic impacts. There were two attributes stemming from the *environmental impact*, *natural TRIX deviation* and *number of birds*. The degree of eutrophication in a coastal area can be expressed as a *TRIX* (TRophic state Index) deviation from the background value. The attribute associated with this lowest-level objective represented the average *TRIX* deviation regarding previous years over a 20-year period.

Another environmental impact we took into account was related to the sharp fall in birddays over the year in recent decades. The associated attribute accounted for the number of birds representing the average number of Bewick's swans and pintails living in the lagoon in a year for the time period under consideration.

Regarding the *social impact* we made a distinction between the *social impact for critical population*, i.e., people living around the lagoon that may be affected by the application of intervention strategies, and *collective social impact*. Both subjective attributes account for aspects like sentiment, possible employment associated with strategy application, crop image…

Finally, the *economic impact* was computed by the average costs concerning the intervention strategy application, i.e., nutrient abatement costs and/or construction and maintenance costs for facilities.

Note that while the models or experts initially provided precise performances, imprecision was introduced by means of an attribute deviation of 10% to evaluate the robustness of the evaluation (see Table 4 in Jiménez et al., 2011).

Next, DM's preferences were quantified accounting for ecocentric, anthropocentric and taxrefuser perspectives, which lead to different weight sets, and, finally, Monte Carlo simulation techniques were applied. Fig. 16 shows the resulting multiple boxplot from both perspectives.

Looking at the multiple box plots for the ecocentric and anthropocentric perspectives, we find that *S*5: *Sluice* and *S*9: *No action* are ranked highest in both boxplots. *S*8: 33% *P abatement + Sluice* is ranked highest from the ecocentric viewpoint, but its best ranking from the anthropocentric perspective is fifth. Finally, *S*6: *Salt7.2*, with a best ranking of second from the anthropocentric viewpoint, is ranked as the worst strategy from the ecocentric perspective. *S*5: *Sluice* and *S*9: *No action* look better than the others. Moreover, the average rankings for both are 1.011 and 2.489 from the ecocentric perspective, respectively, and 1.531 and 1.605 from the anthropocentric viewpoint. These results are even consistent regarding the tax-refuser perspective, in which *S*5 is better ranked (average ranking 1.246) than *S*9

(average ranking 4.621). Thus, we arrived at the conclusion that *S*5: *Sluice* was the intervention strategy to be recommended.
Moreover, if we assume that there is no knowledge whatsoever of the relative importance of the attributes, i.e., weights for the attributes are generated completely at random, *S*5: *Sluice* was again the best intervention strategy throughout the simulation.
The same methodology was applied for different interest rates (0, 2, 4, 6, and 8%), and we arrived at the same conclusion that *S*5: *Sluice* is the intervention strategy to be recommended.

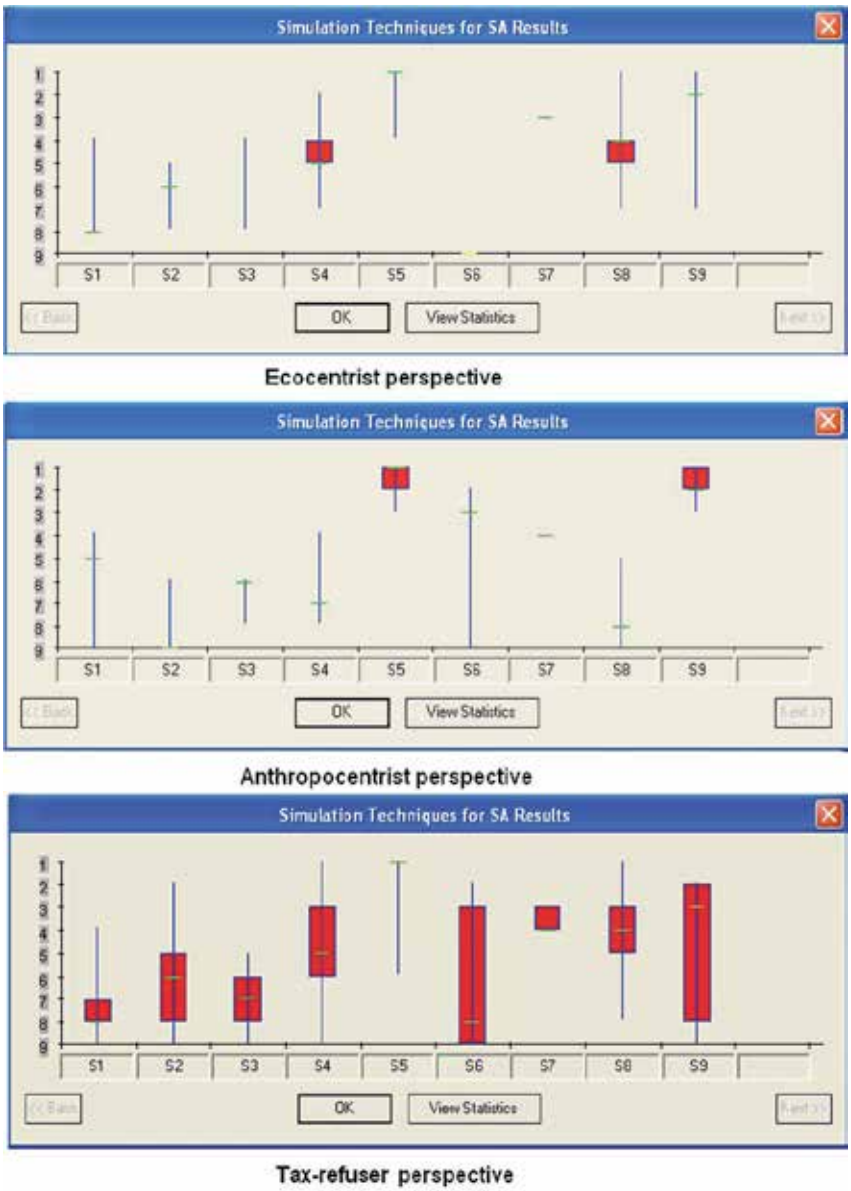


Fig. 16. Strategy evaluation from different perspectives

4. Dominance measuring methods

A recent approach for dealing with incomplete information about weights within *MAUT* is to use information about each alternative's intensity of dominance, known as *dominance measuring methods*.

Let us consider a group decision-making problem with n attributes (X_i, i=1,…,n) and m alternatives (A_j, j=1,…,m), where incomplete information about input parameters has been incorporated into the decision-making process as follows:

- Alternative performances under uncertainty ($x_i^j \in [x_i^{jL}, x_i^{jU}]$, i=1,…,n; j=1,…,m.
- Imprecision concerning utility function assessment ($u_i(\bullet) \in [u_i^L(\bullet), u_i^U(\bullet)]$, i=1,…,n), where $u_i^L(\bullet)$ and $u_i^U(\bullet)$ are the lower and the upper utility functions of the attribute X_i, and
- Imprecision concerning weights, which is represented by weights intervals ($w_i \in [w_i^L, w_i^U]$, i=1,…,n).

Given two alternatives A_k and A_l, alternative A_k dominates A_l if $D'_{kl} \geq 0$, D'_{kl} being the optimum value of the optimization problem (Puerto et al., 2000),

$$\begin{aligned} D'_{kl} = \quad & \min\left\{ u(A_k) - u(A_l) = \sum_{i=1}^{n} w_i u_i(x_i^k) - \sum_{i=1}^{n} w_i u_i(x_i^l) \right\} \\ & w_i^L \leq w_i \leq w_i^U, i = 1,...,n \\ & x_i^{kL} \leq x_i^k \leq x_i^{kU}, i = 1,...,n \\ s.t. \quad & x_i^{lL} \leq x_i^l \leq x_i^{lU}, i = 1,...,n \\ & u_i^L(x_i^{kL}) \leq u_i(x_i^k) \leq u_i^U(x_i^{kU}), i = 1,...,n \\ & u_i^L(x_i^{lL}) \leq u_i(x_i^l) \leq u_i^U(x_i^{lU}), i = 1,...,n \end{aligned} \quad . \tag{2}$$

This concept of dominance is called *pairwise dominance* and leads to the so-called *dominance matrix*:

$$D = \begin{pmatrix} - & D_{11} & \dots & D_{1m-1} & D_{1m} \\ D_{21} & D_{22} & \dots & D_{2m-1} & D_{2m} \\ D_{31} & D_{32} & \dots & D_{3m-1} & D_{3m} \\ & & \dots & & \\ D_{m1} & D_{m2} & \dots & D_{mm-1} & - \end{pmatrix}$$

where $D_{kl} = D'_{kl}/\Sigma w_i$ and D'_{kl} and w_i are the optimum values for the objective function and weights in problem (2), respectively. The dominance matrix is the starting point of dominance measuring methods.

The first two dominance measuring methods were proposed in (Ahn & Park, 2008). In the first, denoted *AP1* (Ahn and Park 1), alternatives are ranked according to a *dominating measure*

$$\alpha_k = \sum_{j=1, j \neq k}^{m} D_{kj}.$$

The higher this dominating measure is the more preferred the alternative will be, because the sum of the intensity of one alternative dominating the others will also be greater. In the

second approach, denoted *AP2* (Ahn and Park 2), alternatives are ranked according to the difference between the dominating measure α_k and a dominated measure

$$\beta_k = \sum_{l=1, l\neq k}^{m} D_{lk},$$

i.e., on the basis of α_k-β_k.

Two new dominance measuring methods were proposed in (Mateos et al, 2011a). The first one, denoted *DME1* (Dominance Measuring Extension 1), is based on the same idea as Ahn and Park suggested. First, *dominating* and *dominated measures* are computed from the paired dominance values and then a *net dominance* is derived. This net dominance is used as a measure of the strength of preference. *DME1* computes the positive and negative dominating measures and positive and negative dominated measures. They are used to compute first a proportion representing how strongly one alternative is preferred to the others and second a proportion representing how intensely one alternative is not preferred to the others. Finally, *DME1* subtracts both proportions to compute the intensity of the preference. *DME1* can be implemented as follows:

1. Get the paired dominance values D_{kl} and the dominance matrix D as in (2).
2. Compute the *dominating measures* α_k, α_k^+ and α_k^- for each alternative A_k:

$$\alpha_k = \sum_{l=1, l\neq k}^{m} D_{kl}, \quad \alpha_k^+ = \sum_{l=1, l\neq k, D_{kl}>0}^{m} D_{kl}, \quad \alpha_k^- = \sum_{l=1, l\neq k, D_{kl}<0}^{m} D_{kl}$$

3. Compute the proportion

$$P_k^{\alpha} = \frac{\alpha_k^+}{\alpha_k^+ - \alpha_k^-}.$$

4. Compute the *dominated measures* β_k, β_k^+ and β_k^- for each alternative A_k:

$$\beta_k = \sum_{l=1, l\neq k}^{m} D_{lk}, \quad \beta_k^+ = \sum_{l=1, l\neq k, D_{lk}>0}^{m} D_{lk}, \quad \beta_k^- = \sum_{l=1, l\neq k, D_{lk}<0}^{m} D_{lk}$$

5. Compute the proportion

$$P_k^{\beta} = \frac{\beta_k^+}{\beta_k^+ - \beta_k^-}.$$

6. Calculate the *preference intensity value* P_k for each A_k: $P_k = P_k^{\alpha} - P_k^{\beta}, k = 1,...,m.$
7. Rank alternatives according to the P_k values, where the best (rank 1) is the alternative for which P_k is a maximum and the worst (rank m) is the alternative for which P_k is the minimum.

The drawback of the *DME1* method is that when the dominance matrix D contains all negative elements, i.e., when all the alternatives are non-dominated, the algorithm is unable to rank the alternatives because they are all equal to 0.

In the second method, denoted *DME2* (Dominance Measuring Extension 2), alternatives are ranked on the basis of a *preference intensity measure*. Paired dominance values are first transformed into *preference intensities* PI_{kl} (step 2) depending on the preference among

alternatives A_k and A_l. Then a *preference intensity measure* (PIM_k) is derived for each alternative A_k (step 3) as the sum of the preference intensities of alternative A_k regarding the others alternatives. This is used as the measure of the strength of preference.
DME2 can be implemented as follows:can be implemented as follows:

1. Compute dominance matrix D from the paired dominance values D'_{kl} (2).
2. If $D_{kl} \geq 0$, then alternative A_k is preferred to alternative A_l, i.e., the intensity with which alternative A_k is preferred to A_l is 1, PI_{kl}=1.
   Else ($D_{kl} < 0$):
   - If $D_{lk} \geq 0$, then alternative A_l dominates alternative A_k, therefore, the intensity with which alternative A_k is preferred to A_l is 0, i.e., PI_{kl}=0.
   - Else note that alternative A_l is preferred to alternative A_k for those values in $\boldsymbol{W_{kl}}$ (constraints of the optimization problem (2)) that satisfy $D_{kl} \leq \Sigma_i\ w_i u_i(x_i^k) - \Sigma_i\ w_i u_i(x_i^l) \leq 0$, and A_k is preferred to A_l for those values in $\boldsymbol{W_{kl}}$ that satisfy $0 \leq \Sigma_i\ w_i u_i(x_i^k) - \Sigma_i\ w_i u_i(x_i^l) \leq -D_{lk} \Rightarrow$ the intensity A_k is preferred to A_l is

$$PI_{kl} = \frac{-D_{lk}}{-D_{lk} - D_{kl}}.$$

3. Compute a preference intensity measure for each alternative A_k

$$PIM_k = \sum_{l=1, l \neq k}^{m} DP_{kl}$$

Rank alternatives according to the *PIM* values, where the best (rank 1) is the alternative with greatest *PIM* and the worst is the alternative with the least *PIM*.
DME1 and *DME2*, like *AP1* and *AP2*, considered ordinal relations regarding attribute weights, i.e., DMs ranked attributes in descending order of importance. For this scenario, Monte Carlo simulation techniques were carried out to analyze their performance and to compare them with other approaches, such as *surrogate weighting methods* (Stillwell et al., 1981; Barron & Barrett, 1996) and adapted *classical decision rules* (Salo & Hämäläinen, 2001).
The results showed that *DME2* performs better in terms of the identification of the best alternative and the overall ranking of alternatives than other dominance measuring methods proposed by different authors. Also, *DME2* outperforms the adaptation of classical decision rules and comes quite close to the r*ank-order centroid weights* method, which was identified as the best approach.
Different cases with incomplete information about weights are considered in (Mateos et al., 2011b). Specifically, we consider weight intervals, weights fitting independent normal probability distributions or weights represented by fuzzy numbers (triangular and trapezoidal). A simulation study was also performed to compare the proposed methods with the measures reported in Ahn and Park, with classical decision rules and with the *SMAA* and *SMAA-2* methods in the above situations. The results show that *DME2* and *SMAA-2* outperform the other methods in terms of the identification of the best alternative and the overall ranking of alternatives.

5. Conclusions

Many complex decision-making problems have multiple conflicting objectives in the sense that further achievement in terms of one objective can occur only at the expense of some

achievement of another objective. Therefore, preference trade-offs between differing degrees of achievement of one objective and another must be taken into account. Also, real problems are usually plagued with uncertainty, and it is impossible to predict with certainty what the consequences of each strategy under consideration will be. Formal analysis is required because it is very difficult to consider the above complexities informally in the mind.

The GMAA *system* is a DSS based on an additive multi-attribute utility model that is intended to allay many of the operational difficulties involved in a decision-making problem. It has been proven through the paper that it very useful to aid DMs in complex decision-making from a different nature, from the restoration of aquatic ecosystems contaminated by radionuclides to the selection of a supplier for cleaning services in an underground transportation company.

On the other hand, a recent approach to deal with imprecise information in multicriteria decision-making are the dominance measuring methods, which has been proved to outperform other approaches, like most surrogate weighting methods or the modification of classical decision rules to encompass an imprecise decision context. In this paper we have reviewed the main dominance measuring methods provided by different authors.

6. Acknowledments

The paper was supported by Madrid Regional Government project S-2009/ESP-1685 and the Spanish Ministry of Education and Science project TIN 2008-06796-C04-02.

7. References

Adelbratt, T. & Montgomery, H. (1980), Attractiveness of Decision Rules, *Acta Psychologica*, Vol. 45, pp. 177-185, ISSN 0001-6918

Ahn, B.S. (2003). Extending Malakooti's Model for Ranking Multicriteria Alternatives with Preference Strength and Partial Information. *IEEE Transactions on Systems, Management and Cybernetics, Part A: Systems and Humans*, Vol. 33, pp. 281-287, ISSN 1083-4427

Ahn, B.S. & Park, K.S. (2008). Comparing Methods for Multiattribute Decision Making with Ordinal Weights. *Computers & Operations Research*, Vol. 35, pp. 1660-1670, ISSN 0305-0548

Ananda, J. & Herath, G. (2009). A Critical Review of Multi-Criteria Decision Making Methods with Special Reference to Forest Management and Planning. *Ecological Economics*, Vol. 68, pp. 2535-2548, ISSN 0921-8009

Bana e Costa, C.A. (1986). Multicriteria Decision Aid Methodology to Deal with Conflicting Situations on the Weights. *European Journal of Operational Research*, Vol. 26, pp. 22-34, ISSN 0377-2217

Bana e Costa, C.A. & Vansnick, J.C. (2008). A Critical Analysis of the Eigenvalue Method used to Derive Priorities in AHP. *European Journal of Operational Research*, Vol. 187, pp. 1422-1428, ISSN 0377-2217

Barron, F. & Barrett, B. (1996). Decision Quality Using Ranked Attribute Weights. *Management Science*, Vol. 42, No. 11, pp. 1515-1523, ISSN 1526-5501

Brans, J.P. & Mareschal, B. (2005). Promethee Methods, In: *Multiple Criteria Decision Analysis. State of the Art Survey*, J.R. Figueira, S. Greco and M. Ehrgott (Eds.), 112-131, Springer, ISBN 038723067X, New York, USA

Brans, J.P. & Vincke, P.H. (1985). A Preference Ranking Organization Method - (the Promethee Method for Multiple Criteria Decision-Making). *Management Science*, Vol. 31, pp. 647-656, ISSN 0025-1909

Brans, J.P.; Vincke, P. & Mareschal, B. (1986). How to Select and How to Rank Projects – the Promethee Method. *European Journal of Operational Research*, Vol. 24, pp. 228-238, ISSN 0377-2217

Bryhn, A.; Jiménez, A.; Mateos, A. & Ríos-Insua, S. (2009). Multi-Attribute Analysis of Trophic State and Waterfowl Management in Ringkøbing Fjord, Denmark. *Journal of Environmental Management*, Vol. 90, No. 8, pp. 2568-2577, ISSN 0301-4797

Bucksaw, D. (2010). Decision Analysis Software Survey. *OR/MS Today*, Vol. 37, No. 6, pp. 46-55, ISSN 1085-1038

Charnetski, J.R. & Soland, R.M. (1978). Multiple-Attribute Decision Making with Partial Information: The Comparative Hypervolume Criterion, *Naval Research Logistics Quaterly*, Vol. 15, pp. 279-288, ISSN 0894-069X

Clemen, R.T. (1996). *Making Hard Decisions*, Duxbury Press, ISBN 0-534-92336-4, Belmont, USA

Corner, J.L. & Kirkwood, C.W. (1991). Decision Analysis Applications in the Operations Research Literature, 1970-1989. *Operations Research*, Vol. 39, pp. 206-219, ISSN 1526-5463

Corsair, H.J.; Ruch, J.B.; Zheng, P.Q.; Hobbs, B.F. & Koonce, J.F. (2009). Multicriteria Decision Analysis of Stream Restoration: Potential and Examples. *Group Decision and Negotiation*, Vol. 18, pp. 387-417, ISSN 0926-2644

Edwards, W. (1977). How to Use Multiattribute Utility Measurement for Social Decision-Making. *IEEE Transactions on Systems, Managements and Cybernetics*, Vol. 7, No. 5, pp. 326-340, ISSN 1083-4427

Edwards, W. & Barron, F.H. (1994). SMART and SMARTER: Inproved Simple Methods for Multiattribute Utility Measurement. *Behaviour and Human Decision Processes*, Vol. 60, pp. 306-325, ISSN 0749-5978

Eum, Y.; Park, K.S., & Kim, H. (2001). Establishing Dominance and Potential Optimality in Multicriteria Analysis with Imprecise Weights and Values. *Computers and Operations Research*, Vol. 28, pp. 397-409, ISSN 0305-0548

Figueira, J.; Greco, S. & Ehrgott, M. (eds.) (2005). Multiple Criteria Decision Analysis. State of the Art Surveys, Springer, ISBN 0-387-23067-X, New York, USA

Figueira, J.; Mousseau, V. & Roy, B. (2005). *Electre Methods,* In: *Multiple Criteria Decision Analysis. State of the Art Survey*, J.R. Figueira, S. Greco and M. Ehrgott (Eds.), 133-162, Springer, ISBN 038723067X, New York, USA

Geneletti, D. (2005). Formalising Expert Opinion Through Multi-Attribute Value Functions: An Application in Landscape Ecology. *Journal of Environmental Management*, Vol. 76, pp. 255-262, ISSN 0301-4797

Håkanson, L., & Bryhn, A.C. (2008). Goals and Remedial Strategies for Water Quality and Wildlife Management in a Coastal Lagoon – A Case-Study of Ringkøbing Fjord, Denmark. *Journal of Environmental Management*, Vol. 86, pp. 498–519, ISSN 0301-4797.

Håkanson, L.; Bryhn, A.C.& Eklund, J.M. (2007). Modelling Phosphorus and Suspended Particulate Matter in Ringkøbing Fjord in Order to Understand Regime Shifts. *Journal of Marine Systems*, Vol. 68, pp. 65–90, ISSN 0924-79-63.

Hazen, G.B. (1986). Partial Information, Dominance and Potential Optimality in Multiattribute Utility Theory. *Operations Research*, Vol. 34, No. 2, pp. 296-310, ISSN 0030-364X

Ito, K.; Xu, Z.X.; Jinno, K.; Hojiri, T. & Kawamura, A. (2001). Decision Support System for Surface Water Planning in River Basins. *Journal of Water Resources Planning and Management,* Vol. 127, pp. 272–276, ISSN 0733-9496

Jiménez. A., Ríos-Insua, S. & Mateos, A. (2003). A Decision Support System for Multiattribute Utility Evaluation Based on Imprecise Assignments. *Decision Support Systems*, Vol. 36, No. 1, pp. 65–79, ISSN 0167-9236

Jiménez, A.; Rodríguez, L.C.; Ríos-Insua, S. & Mateos, A. (2005). Selection of Segments to be Sourced from Low Cost Countries for a Global Industrial Equipment Manufacturer based on a Multi-Attribute Decision Support System. *Research in Computing Science,* Vol. 17, pp. 3-12, ISSN 1665-9899

Jiménez, A.; Ríos-Insua, S. & Mateos, A. (2006). A Generic Multi-Attribute Analysis System. *Computers & Operations Research,* Vol. 33, No. 4, pp. 1081-1101, ISSN 0305-0548

Jiménez, A.; Mateos, A.; Ríos-Insua, S. & Rodríguez, L.C. (2007). Contracting of Cleaning Services in a European Underground Transportation Company with the Aid of a DSS. *Decision Support Systems,* Vol. 43, No. 4, pp. 1485-1498, ISSN 0167-9236

Jiménez, A.; Mateos, A. & Brhyn, A. (2011). Selecting Intervention Strategies Against Eutrophication and the Drastic Decrease in Bird Abundance in Ringkøbing Fjord. *TOP An Official Journal of the Spanish Society of Statistics and Operations Research* (accepted, DOI: 10.1007/s11750-010-0136-x), ISSN 1134-5764

Joubert, A.; Stewart, T.J. & Eberhard, R. (2003). Evaluation of Water Supply Augmentation and Water Demand Management Options for the City of Cape Town. *Journal of Multi-Criteria Decision Analysis*, Vol. 12, pp. 17-25, ISSN 1057-9214

Keefer, D.L.; Kirkwood, C.W. & Corner, J.L. (2004). Perspective on *Decision Analysis* Applications, 1990-2001. *Decision Analysis*, Vol. 1, pp. 4-22, ISSN 1545-8490

Keeney, R.L. & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, ISBN 0521438837, Cambridge, UK

Kirkwood, C.W. (1997). *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*, Duxbury, ISBN 0-534-51692-0, Beltmon, USA

Kirkwood, C.W. & Sarin, R.K. (1985). Ranking with Partial Information: A Method and an Application. *Operations Research*, Vol. 33, pp. 38-48, ISSN 0030-364X

Kmietowcz, Z.W. & Pearman, A.D. (1984). Decision Theory, Linear Partial Information and Statistical Dominance. *Omega,* Vol. 12, pp. 391-399, ISSN 0305-0483

Lahdelma, R. & Salminen, P. (1998). SMAA - Stochastic Multiobjective Acceptability Analysis. *European Journal of Operational Research*, Vol. 106, pp. 137-143, ISSN 0377-2217

Lahdelma, R. & Salminen, P. (2001). SMAA2 - Stochastic Multiobjective Acceptability Analysis for Group Decision Making. *Operations Research*, Vol. 49, pp. 444-454, ISSN 0030-364X

Lee, K.; Park, K.S. & Kim, H. (2002). Dominance, Potential Optimality, Imprecise Information, and Hierarchical Structure in Multi-criteria Analysis. *Computers and Operations Research*, Vol. 29, pp. 1267-1281, ISSN 0305-0548

Linkov, I.; Satterstrom, F.K.; Kiker, G.; Seager, T.P.; Bridges, T.; Gardner, K.H., Rogers, S.H.; Belluck, D.A. & Meyer, A. (2006). Multicriteria Decision Analysis: A Comprehensive Decision Approach for Management of Contaminated Sediments. *Risk Analysis*, Vol. 26, pp. 61-78, ISSN 0272-4332

Malakooti, B. (2000). Ranking and Screening Multiple Criteria Alternatives with Partial Information and Use of Ordinal and Cardinal Strength of Preferences. *IEEE Transactions on Systems, Management and Cybernetics, Part A: Systems and Humans*, Vol. 30, pp. 787-801, ISSN 1083-4427

Mateos, A.; Jiménez, A. & Blanco, J.F. (2011a). A MCDM Ranking Method Based on a Dominance Measure: Computational Study. *Group Decision and Negotiation* (submitted), ISSN 0926-2644

Mateos, A.; Jiménez, A. & Blanco, J.F. (2011b). Dominance Measuring Method Performance under Incomplete Information about Weights. *Journal of Multi-Criteria Decision Analysis* (submitted), ISSN 1057-9214

Mateos, A.; Jiménez, A. & Ríos-Insua, S. (2006). Monte Carlo Simulation Techniques for Group Decision-Making with Incomplete Information. *European Journal of Operations Research,* Vol. 174, No. 3, pp. 1842-1864, ISSN 0377-2217

Mateos, A.; Ríos-Insua, S. & Jiménez, A. (2007). Dominance, Potential Optimality and Alternative Ranking in Imprecise Decision Making. *Journal of Operational Research Society,* Vol. 58, No. 3, pp. 326-336, ISSN 0160-5682

Mendoza, G.A. & Martins, H. (2006). Multi-Criteria Decision Analysis in Natural Resource Management: A Critical Review of Methods and New Modelling Paradigms. *Forest Ecology and Management*, Vol. 230, pp. 1-22, ISSN 0378-1127

Merkhofer, M.W.; Conway, R. & Anderson, R.G. (1997). Multiattribute Utility Analysis as a Framework for Public Participation in Siting a Hazardous Waste Management Facility. *Environmental Management*, Vol. 21, pp. 831-839, ISSN 0044-7447

Monte, L.; Kozhouharov, V.; Zheleznyak, M.; Kryshev, I.; Voitsekhovitch, O.; Brittain, J.; Gallego, E. & Hakanson, L. (eds.) (2002). Implementing Computerised Methodologies to Evaluate the Effectiveness of Countermeasures for Restoring Radionuclide Contaminated Fresh Water Ecosystems. COMETES Project. *Final Report RT/AMB/2001/2, Serie Ambiente*, ENEA, Rome. ISSN 1120-5555

Monte, L.; van der Steen, J.; Bergstroem, U.; Gallego, E.; Brittain, J. & Hakanson, L. (eds.) (2000). The Project MOIRA: A Model-Based Computerised System for Management Support to Identify Optimal Remedial Strategies for Restoring Radionuclide Contaminated Aquatic Ecosystems and Drainage Areas. *Final Report RT/AMB/2000/13, Serie Ambiente*, ENEA, Rome. ISSN 1120-5555

Nijkamp, P.; Rietveld, P. & Voogd, H. (1990). *Multicriteria Evaluation Physical Planning*, North Holland, ISBN 0444881247, Amsterdam, The Netherlands

Park, K. (2004). Mathematical Programming Models for Characterizing Dominance and Potential Optimality when Multicriteria Alternative Values and Weights are Simultaneously Incomplete. *IEEE Transactions on Systems, Management and Cybernetics, Part A: Systems and Humans*, Vol. 34, pp. 601-614, ISSN 1083-4427

Perez, J.; Jimeno, J.L. & Mokotoff, E. (2006). Another Potential Shortcoming of AHP. *TOP An Official Journal of the Spanish Society of Statistics and Operations Research*, Vol. 14, pp. 99-111, ISSN 1134-5764

Petersen, J.K.; Hansen, J.W.; Brogaard Laursen, M.; Clausen, P.; Carstensen, J. & Conley, D.J. (2008). Regime Shift in a Coastal Marine Ecosystem. *Ecological Applications,* Vol. 18, pp. 497–510, ISSN 1051-0761

Poch, M., Comas, J., Rodríguez-Roda, I.; Sánchez-Marré, M. & Cortés, U. (2004). Designing and Building Real Environmental Decision Support Systems. *Environmental Modeling & Software,* Vol. 19, No. 9, pp. 857-873, ISSN 1364 -8152

Prato, T. & Herath, G. (2007). Multiple-Criteria Decision Analysis for Integrated Catchment Management. *Ecological Economics*, Vol. 63, pp. 627-632, ISSN 0921 -8009

Puerto, J.; Marmol, A.M.; Monroy, L. & Fernández, F.R. (2000). Decision Criteria with Partial Information. *International Transactions in Operational Research*, Vol. 7, pp. 51-65, ISSN 1475-3995

Raiffa, H. (1982). *The Art and Science of Negotiation*. Harvard University Press, ISBN 0-674-04813, Cambridge, UK

Reichert, P.; Borsuk, M.; Hostmann, M.; Schweizer, S.; Sporri, C.; Tockner, K. & Truffer, B. (2007). Concepts of Decision Support for River Rehabilitation. *Environmental Modelling & Software*, Vol. 22, pp. 188-201, ISSN 1364-8152

Ríos Insua, D. & French, S. (1991). A Framework for Sensitivity Analysisin Discrete Multi-Objective Decision-Making. *European Journal of Operational Research*, Vol. 54, pp. 176-190, ISSN 0377-2217

Ríos-Insua, S.; Jiménez, A. & Mateos, A. (2004). A Time-Dependent Decision Support System for MultiAttribute Decision-Making. *Integrated Computer-Aided Engineering,* Vol. 11, No. 1, pp. 63-75, ISSN 1069-2509

Ríos-Insua, S.; Gallego, E.; Jiménez, A. & Mateos, A. (2006). A Multi-Attribute Decision Support System for Selecting Environmental Intervention Strategies for Radionuclide Contaminated Freshwater Ecosystems. *Ecological Modelling,* Vol. 196, No. 1-2, pp. 195-208, ISSN 0304-3800

Roy, B. (1996). *Multicriteria Methodology for Decision Aiding*. Kluwer Academic Publishers, ISBN 079234166X, Dordrecht, The Netherlands

Saaty, T.L. (1980). *The Analytic Hierarchy Process*, MacGraw-Hill, ISBN 0070543712, New York, USA

Sage, A. & White, C.C. (1984). Ariadne: a Knowledge-Based Interactive System for Planning and Decision Support. *IEEE Transactions on Systems, Management and Cybernetics, Part A: Systems and Humans*, Vol. 14, pp. 35-47, ISSN 1083-4427

Salo, A.A. & Hämäläinen, R.P. (1995). Preference Programming through Approximate Ratio Comparisons. *European Journal of Operations Research*, Vol. 82, pp. 458-475, ISSN 0377-2217

Salo, A.& Hämäläinen, R.P. (2001). Preference Ratio in Multiattribute Evaluation (PRIME) - Elicitation and Decision Procedures under Incomplete Information. *IEEE Transactions on Systems, Management and Cybernetics, Part A: Systems and Humans,* Vol. 31, No.6, pp. 533-545, ISSN 1083-4427

Sarabando, P. & Dias, L.C. (2010). Simple Procedures of Choice in Multicriteria Problems Without Precise Information About the Alternatives Values. *Computers & Operations Research*, Vol. 37, No. 12, pp. 2239-2247, ISSN 0305-0548

Stam, A.; Salewicz, K.A. & Aronson, J.E. (1998). An Interactive Reservoir System for Lake Kariba. *European Journal of Operational Research*, Vol. 107, pp. 119-136, ISSN 0377-2217

Stewart, T.J. (1996). Robustness of Additive Value Function method in MCDM. *Journal of Multi-Criteria Decision Analysis*, Vol. 5, pp. 301–309, ISSN 1099-1360

Stillwell, W.G.; Seaver, D.A. & Edwards, W.A. (1981). Comparison of Weight Approximation Techniques in Multiattribute Utility Decision Making. *Organization Behavior and Human Decision Processes*, Vol. 28, pp. 62-77, ISSN 0749-5978

Tecle, A.; Shrestha, B.P. & Duckstein, L. (1998). A Multiobjective Decision Support System for Multiresource Forest Management. *Group Decision and Negotiation*, Vol. 7, pp. 23-40, ISSN 0926-2644

Voogd, H. (1983). *Multicriteria Evaluation for Urban and Regional Planning*, Pion Ltd., London, UK.

Wang, X. & Triantaphyllou, E. (2008). Ranking Irregularities when Evaluating Alternatives by Using some ELECTRE Methods. *Omega*, Vol. 36, pp. 45-63, ISSN 0305-0483.

Weber, M. (1987). Decision Making with Incomplete Information. *European Journal of Operational Research*, Vol. 28, pp. 44-57, ISSN 0377-2217

von Winterfeldt, D. & Edwards, W (1986), *Decision Analysis and Behavioral Research*, Cambridge University Press, ISBN 0521273048, Cambridge, UK

Zeithaml, V.A.; Parasuraman, A. & Berry, L.L. (1990). *Delivering Quality Service: Balancing Customer Expectations and Perceptions*, The Free Press, ISBN 9780029357019, New York, USA.

Part 3

Multidisciplinary Case Studies

15

Decision Support Systems in Agriculture: Some Successes and a Bright Future

Russell Yost, Tasnee Attanandana, Carol J. Pierce Colfer and Stephen Itoga
University of Hawai`i at Manoa
USA

1. Introduction

This chapter is dedicated to the capture, preservation, reuse and learning of agricultural knowledge. We illustrate the potential of information technology with a simple example of a writing pen. Before the age of information, a person with a pen or pencil could write beautiful poetry. If she gave or loaned the pen to someone else, she could no longer write poetry, but the other person, would have gained a tool that helped write poetry. In this case, the first person's loss is the second person's gain, which, in economic terms, is a zero-sum game. Of importance is the fact that the relationship between the first and second person has changed and in order to continue writing poetry the first must obtain permission from the second to continue writing. Thus a certain measure of power has been passed with the possession of the pen and a dependency has changed between the first person and the second. Rare is the individual that does not see this as a clear disadvantage for the first person. Also rare is the relationship between two people that would not be strained by such a reversal.

Imagine, however, if the first person were to make a copy of the pen and give it to the second person, while retaining the use of the pen and thus suffer no loss in ability to write poetry. The relationship between the first person and the second changes from one of dependency to one of collaboration and mutual empowerment. Rare is the relationship between two persons that would not be strengthened rather than strained by the sharing of our information age pen. In this case no longer is it a zero-sum transaction. No longer is there the capable and the incapable. No longer is there gain of one at the loss of the other. Rather it has become a win-win situation in which all gain. Under these conditions, the first person is more likely and could be stimulated to make copies and distribute pens to everyone in the world, since it no longer results in their losing the tools to write poetry. This is the potential of information technology. Information technology can enable and empower us to share tools without the loss of use of the tool ourselves. It seems we have yet to fully exploit the potential of this technology.

2. Scope of this chapter

We will concentrate this chapter on agricultural knowledge, particularly that pertinent and relevant to tropical agroecosystems, largely because the bulk of our experience with decision-aids has been concerned with such production systems. Our thesis is that successful decision-aids need to recognize the inherent complexity of such systems. It is the thesis of this chapter that decision-aids can be tools to assist in the management of these

complex yet critical elements of human food security, partially through the capture of relevant knowledge and also through facilitating the accelerated learning/acquisition of the knowledge by others, and also through improvement in that knowledge as a result of the organization and representation effort.

This chapter will describe some of the authors' experience with decision-aids and their characteristics that have been useful in agriculture. The initial motivation to develop a decision-aid derived from the confluence of four conditions occurring at the onset of a newly formed, foreign technical assistance project in Indonesia (TropSoils, 1981):

1. The goal of the project was to provide improved soil and crop management for Transmigrants (farmers and producers from the "over-populated" rural areas of Java, Indonesia) in their new environment on Sumatra with relatively large amounts of land, but little land with the water needed for paddy rice cultivation, the system anticipated by government planners and desired by some farmers. The new homesteads in Sumatra provided little land suitable for paddy rice production. The more extensive land differed drastically from that on Java by being exceedingly acid, with pH values of 4.0 to 4.5 and high levels of plant toxic aluminum. Aluminum saturation values frequently exceeded 50%, indicating probable toxicity to food crop plants such as maize (*Zea mays*, L), peanut (*Arachis hypogea*, L.), and especially mung bean (*Vigna radiata*). Other soil constraints to food crop productivity included low levels of essential soil nutrients (phosphorus, potassium), which also were constraints rare in the rich Javanese soils. Thus the need was great to provide new ways for the Transmigrants to produce food and secure a livelihood in this strange, new environment.
2. Two US universities were tapped to provide the technical assistance (North Carolina State University, and the University of Hawai`i at Manoa). These universities had extensive experience dealing with soils taxonomically identical (Paleudults[1]) to those at the project site (Sitiung, West Sumatra). The immediate challenge was "How could the experience of producers and growers in the SouthEast US, Central and South America, which was largely experiential, but also recently scientific, be efficiently introduced and shared with the Transmigrants," who were in immediate need of food production technology on their new, but unfamiliar land.
3. A new perspective had just appeared in international agricultural development research circles, that of Farming Systems Research and Development (Shaner et al., 1982). The approach pointed out that farmers should be respected and very much involved in attempts to introduce new technology and practice. This approach also seemed to coalesce with Agroecosystems Analysis, as advocated by South East Asian scientists in the SUAN network (Rambo and Sajise, 1984).
4. Recent developments in information technology, specifically the new capabilities of software development efforts associated with Artificial Intelligence (Rich, 1983), were purported to permit medical diagnosis (Hayes-Roth et al., 1983). It was hypothesized at the time that the detection and possibly the prescription of soil and crop management solutions to the weathered, acid soils, would be analogous to the diagnosis and prescription of appropriate medication in similarly complex human health situations.

With this motivation the initial decision-aids were developed with the perhaps pompous title of "expert systems"(Yost et al., 1988).

---

[1] Paleudults are soils of the Ultisol order, which are particularly old and highly weathered, associated with high, usually year-long rainfall. See Buol, S.W., F.D. Hole, and R.J. McCracken. 1989. Soil Genesis and Classification. 3rd ed. Iowa State University, Ames.

Because decision-aids are often developed to improve the capture, transfer (in a learning sense), and use of agricultural knowledge, the search for and development of successful decision-aids needs to begin with a thorough knowledge of agriculture. It might be yet more appropriate to search for the agricultural knowledge that is critical for providing food security and well-being. One of the hypotheses of the decision-aids effort described herein was that it was possible to capture, transfer (in a learning sense), and use this knowledge more directly---in contrast to simply writing chapters, books, and articles on the knowledge which must then be read and assimilated before the knowledge could be used.

3. Agricultural knowledge

Agricultural knowledge, that is the combined experience of how to grow and produce food, fiber, and bioproducts while securing a livelihood from the land, is extremely complex, comprised of multiple disciplines, multiple persons, with multiple levels of abstraction. Producer decisions range from considering the details of how a plant needs protection from pests and diseases, to planning commodity trading and marketing---all sometimes in a matter of minutes. Other producers' worries range from which variety of food crop to plant, to which field to plant first, to issues of food availability and alternative sources of income should food production fail. With such complexity, uncertainty, and variation over time, it is not surprising that agriculture as an enterprise is considered highly risky. White (personal communication, Cornell University, 2011) clusters the modern agricultural risks into 5 groups (1) Production risk, 2) Marketing/Price Risks, 3) Financial Risk, 4) Legal and Environmental Risk, 5) Human Resource Risks. Of those risks the primary one to be considered in this chapter is production risk. Production risk or productivity has been identified as an agroecosystem property by Conway (1986). He groups the major properties of agroecosystems thusly: 1) Productivity, 2) Stability, and 3) Resilience. Rambo and Sajise (1984) have expanded the number of properties to include those related to the human community associated with the agro-ecosystem (Table 1).

Property	Description
Productivity	The agroecosystem's output of goods and services.
Stability	The degree to which productivity remains constant.
Sustainability (now Resilience)	The ability of a system to maintain its productivity when subjected to stress and shock.
Equitability	A measure of how evenly the products of the agroecosystem are distributed among its human beneficiaries.
Autonomy	A measure of the extent to which a system's survival is dependent on other systems outside its control.
Solidarity	The ability of a social system (i.e. community) to make and implement decisions about collective behavior.
Diversity (Rambo, 1989)	Measure of the number of different kinds/types of components. Usually providing a greater range of options for change when necessary.
Adaptability (Rambo, 1989)	The ability of the system to respond to change in its environment to ensure continuing survival.

An example analysis of several agricultural systems from this perspective is given in (Yost et al., 1997).

Table 1. Agroecosystem properties. (Conway, 1987, Marten and Rambo, 1988).

4. Experiential knowledge

While the totality of agricultural knowledge is, as indicated above, exceedingly complex and diverse, we will consider a small subset of that knowledge in this chapter. We will focus on knowledge related to the growth and production of agricultural food crops and the role of nutrients, either in deficit or excess in that relationship. Agricultural knowledge is extremely descriptive with many adjectives and nouns, but few of the axioms, postulates, and theorems enjoyed by sciences such as physics and mathematics. Also as suggested above, agricultural knowledge tends to be encyclopedic with relatively few universal, nearly inviolable rules. In addition to exercising relatively few universal rules it is also clearly interdisciplinary, requiring close interaction among disciplines to adequately capture the experience.

Acknowledging the interdisciplinarity is important because the methods and norms of the various disciplines differ and should be respected in order to obtain the best knowledge from each of the disciplines. A personal experience illustrates differences among social and biological scientists, for example. Among biological scientists data almost always refers exclusively to numerical knowledge, weights of maize, metric tons of root crops, dollars per kilogram, kilograms of fertilizers or amendments, duration of crop cycles, while social science data can be notes taken during an intensive interview, during a focus group discussion, or as a result of a recollection. It is important in working with such diverse, interdisciplinary knowledge that disciplines are respected for their methods, techniques, approaches and culture.

4.1 Collecting and recording agricultural knowledge

Accurate collection and recording of agricultural knowledge, not surprisingly, must reflect the complexity of the knowledge itself. Such collection is difficult and success, not surprisingly, seems to require methods appropriate for the knowledge. Probably some of the best methods from the point of view of completeness are those used by anthropologists. Their holistic perspective requires unusually complete, thorough knowledge collection and recording using the most current methods available. One good example is the Ph.D. dissertation of Dr. Cynthia T. Fowler (Fowler, 1999), describing an agricultural community, Kodi, West Sumba, Indonesia. The dissertation required approximately 550 pages to record the relevant knowledge. A small portion of the dissertation was later synthesized into an explanation of an apparent oddity – that an introduced plant from another continent came to be a local 'sacred' plant (Fowler, 2005).

Another example of the capture of detailed agricultural knowledge is provided by the dissertation of Dr. M. Robotham (Robotham, 1998). Again, some 550 pages were needed to describe the agricultural system. In this case, Robotham attempted to generalize the knowledge and capture the decision-making logic from each of three villages located in the Philippines (ibid, 1998). Within each of the 3 sites, selected to represent variation in Philippine agriculture, multiple households were interviewed using social science techniques, with a total of some 17 households interviewed in all. Models of the apparent decision-making process were synthesized into decision-trees (graphs that represent the flow of decision-making, Appendix 1) to help compare and contrast the knowledge that had been developed for each of the villages.

Influences of socio-economic forces on agroforestry adoption in the Dominican Republic were modeled using a rule-based system (Robotham, 1996). Examples of a rule-based system will be forthcoming in section 5.2

Another effort, conducted by members of the TropSoil team of our study in Sumatra, Indonesia, was the attempt to capture the similarities and differences among the local people in contrast with the Transmigrants (Colfer et al., 1989). The results suggested that the rule-based knowledge representation structure was not ideal to capture and structure the information. It may have been that the knowledge was descriptive while the software was designed to capture decisions built on goals and rule-based logic.

5. Contributions of artificial intelligence to decision-aid development

(Rich, 1983) defines artificial intelligence (AI) as "the study of how to make computers do things at which, at the moment, people do better." She goes on to list various topics of interest ("problems") as of the time of her book that scientists in the field were working on:

- Knowledge representation
- Search strategies
- Reasoning methods
- Game playing
- Theorem proving
- General problem solving
- Perception (visual, speech)
- Natural language understanding
- Expert problem solving (Symbolic mathematics, Medical diagnosis, Chemical analysis, Engineering design)

Of particular interest to the authors of this chapter was the type of "Expert problem solving" of Medical diagnosis. This application of A.I. illustrates three contributions of A.I. to agricultural knowledge: *Knowledge representation*, *Search Strategies*, and *Reasoning Methods*.

5.1 Characteristics of experts

Glaser and Chi (1988) suggest that experts often display the following characteristics

- Excel mainly in their own domains
- Perceive large meaningful patterns in their domain
- Work quickly. They are faster than novices in performing the skills of their domain
- Have superior short term and long term memory
- See and represent a problem in their domain at a deeper (more principled) level than novices
- Spend a great deal of time analyzing a problem qualitatively
- Have strong self-monitoring skills

5.2 Knowledge representation

One of the first systems to carry out medical diagnosis was the software Mycin (Hayes-Roth et al., 1983) , which used a rule-based system to record and exercise expert knowledge. Rule-based systems were constructed from a sequence of "if then" statements illustrated as follows:

1. If Blood temperature is warm and method of reproduction is live

Then Animal = mammal

2. If Blood temperature is warm and method of reproduction is eggs

Then Animal = bird

The analogy seems obvious between diagnosing and solving a medical condition and that of diagnosing and solving a condition that is constraining or limiting a plant or food crop. This analogy was first recognized by several plant pathologists and resulted in the development of a system to detect soybean diseases (Michalski et al., 1981).
This structure was used in the first 'expert systems' developed by the authors. Rules used to capture the knowledge included, for example:
Rule 1: If the plant observed in the field is *Leucaena leucocephala,* L. and the plant is growing well then it is very unlikely that soil acidity would limit most food crop yields (80/100).
Rule 2: If the soil of the field is red and in a tropical environment then it is likely that soil acidity will limit food crop yields (60/100).
Rules 1 and 2 illustrate ways that observational information, i.e. the presence of a particular plant, can be recorded and can contribute to a conclusion that soil acidity may or may not be limiting. Rule-based systems were used to develop a wide range of diagnostic systems. In addition, these two rules illustrate a method not only to capture the logic in the if-then sequence, but also record some expression of uncertainty in the declaration of the logical relationship. In advanced rule-based systems combinations or rules with less than 100% confidence level would be combined to represent that uncertainty in the resulting conclusion. Some scientists developed methods of checking the consistency of combinations of various rules, by examining the veracity of the resulting conclusion.
Other methods of knowledge representation have been developed such as frames, semantic nets, but these are beyond the scope of this chapter. Given the complexity of agricultural knowledge, improvements in structures supporting knowledge representation continue to be needed. Specifically challenging are ways to combine qualitative and quantitative knowledge in ways that conserve both. Unfortunately, many combinations of these types of knowledge are possible only when the quantitative information is simplified to match the form of the qualitative and when the qualitative is expressed only in quantitative terms.

5.3 Search strategies

As indicated in Rich (1983) and other references, strategies for efficient search through huge networks, decision-trees and databases are needed. AI has provided some clear examples of search strategies such as a) Depth-first, b)Breadth-first, and 3)Best-first (Figure 1). A Depth-first strategy probes a knowledge-tree or a decision-tree by asking the detailed questions first in a limb of the tree (top downward) as the first path through the tree. A Breadth-first strategy, in contrast, searches all nodes at the same depth and then proceeds to the next lower level of nodes (or questions). The Best-first, however, is a combination of the best features of the Depth-first and the Breadth-first strategy. The Best features are those in which a heuristic[2], or specific knowledge, guides the search to choose at each node either a Depth-first or a Breadth-first strategy, depending on the knowledge. It's interesting to note that novices often choose a Depth-first strategy in probing a decision-tree and sometimes ask far too-detailed questions (deep into the decision-tree) too quickly, resulting in a failed search. In fact, this occurs so often that when someone exercises a Depth-first search, and it

[2] On a personal note, my Brazilian wife has shown me a very practical 'heuristic', she taught me how to cook rice by using the rule 1) add rice to the pan and 2) add only enough water to cover the rice by the depth of the distance between the tip of one's index finger and the first joint. Interestingly, some have speculated that this distance may coincide with the "inch" in English measurements. Less controversial is that this as an extremely convenient meter stick!

fails to find the correct answer, we tend to conclude that person is a novice! Experts frequently use the Best-first, where they may switch between search strategies based on their experience and awareness of the decision-tree content.

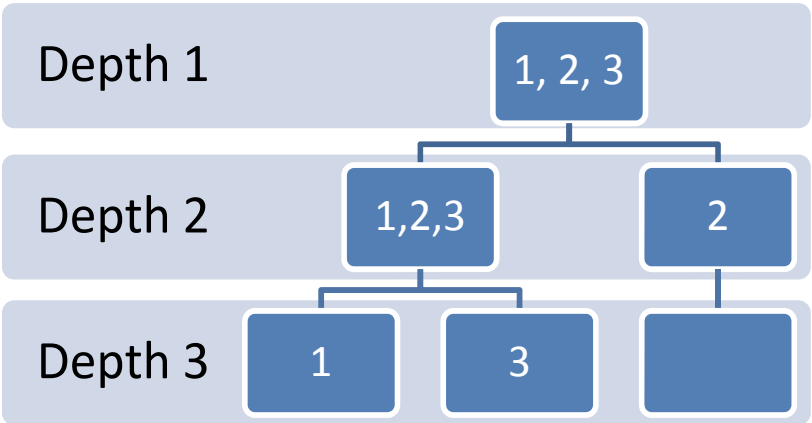


Fig. 1. Decision-tree illustrating Depth-first searches (pathway example 1), Breadth-first (follow pathway 2), and Best-first (follow pathway 3).

Recently, there has been renewed interest in search strategies that can exploit the rapidly expanding information base on the Internet (Watson-Jeopardy, 2011). These strategies may make qualitative information much more accessible to computer based reasoning systems.

5.4 Reasoning methods

A third contribution of AI to agricultural decision-aids (the first being *Knowledge representation*, the second is *Search Strategies*) is the choice between forward-chaining and backward chaining in terms of flow of the reasoning or inference through the decision-tree or set of rules. The forward-chaining method of reasoning begins with the observed facts and makes all possible inferences on the first pass through the decision-tree. The second and subsequent passes collect all facts originally observed plus all conclusions resulting from the first pass through the decision-tree. When the entire decision-tree is evaluated and all possible inferences are made the process is complete. The backward-chaining method begins with the same decision-tree but first evaluates the "goals" or final conclusions or inferences of the decision-tree, of which there typically only a few. Each of these "goals" is evaluated one at a time by determining what facts are needed for each of the goals to be concluded (succeed in being inferred). If any facts are missing that are needed for a specific goal, then that goal is discarded and the next unevaluated goal is similarly evaluated. Many of the initial expert system software programs chose backward-chaining as a reasoning strategy. The backward-chaining method of reasoning or progress through the decision-tree is often much more rapid than forward-chaining because major portions of the decision-tree are truncated if any rule does not have all of the necessary information and thus is evaluated as false. Readers interested in further details of these reasoning strategies are encouraged to consult recent texts or summaries on AI.

As this chapter is being written, new techniques of reasoning are illustrating that machines such as IBM's Watson can account for uncertainty in information and situations, by rank ordering multiple solutions to a given problem. The result is better performance than the best human players of the game Jeopardy (Watson-Jeopardy, 2011). This event is sure to be a

milestone in terms of managing complex and uncertain information, far exceeding the previous success of IBM's Deep Blue that excelled with the much more algorithmic game of chess. The success by Watson included the access of 15 terabytes of information accessed by 10 racks of IBM Power 750 servers, which generated 80 teraflops of processing power. Henshen (2011) reports that the 80 teraflops of processing power together with improved information access methods reduced information access time from 2 hours to 3 seconds.

6. Example decision-aids

Some of the authors used an expert system development tool (expert system shell) to implement a rule-based system that used backward-chaining to diagnose acid soil conditions and prepare predictions of the amount of agricultural limestone needed to remove the soil acidity limitation to selected crops. This software, ACID4, was described in (Yost et al., 1986) and subsequent decision-aids. We now present a list of various decision-aids developed and illustrate the range of uses, methods of implementation, purposes as well as unexpected benefits.

6.1 ACID4 rule-based system

6.1.1 Goal

Facilitate the transfer of the acid soil management knowledgebase developed in Southeast US, Central and South America to farmers and producers of the Transmigration area of Indonesia in general and Sumatra in particular.

6.1.2 Objectives

Implement a set of rules that together represent both the scientific knowledge and farmer experience in managing acid soils for food crop production. The primary source for the knowledge was a review paper by (Kamprath, 1984), practical experience reported by (Gonzalez-Erico et al., 1979), and firsthand experience by the authors.

6.1.3 Implementing language

EXSYS, expert system "shell"(Hunington, 1985.)

6.1.4 Successes

The ACID4 decision-aid illustrated that soil management knowledge could, indeed, be captured and represented in a computer-based decision-aid. The system permitted non-experts with only inputs of measured soil acidity (KCl-extractable acidity, calcium and magnesium and a selected crop) to receive predictions of lime requirements in tons of limestone per hectare (eq. 1).

$$\begin{aligned}&\text{Lime requirement}\left(\text{tons / hectare}\right) = \\ &1.4\left(\text{Exchangeable Acidity} - \text{CAS} * \text{ECEC}\right) / 100\end{aligned} \tag{1}$$

- Where: Lime requirement is the amount of limestone of 100% $CaCO_3$ quality,
- Exchangeable Acidity is the KCl-extractable toxic aluminum and hydrogen,
- CAS is the Critical Aluminum Saturation, which is the maximum amount of toxic aluminum and hydrogen the specific crop can tolerate while achieving maximum yields.

- ECEC is the soil effective cation exchange capacity, which is the sum of the cations (Al, Ca, Mg, and K) as measured by a neutral salt.

The predictions of amount of limestone thus included current soil status of soil acidity, crop needs, quality of the limestone, and ancillary needs of calcium and magnesium in the soil.

An extensive comparison of ADSS, which was a slightly improved version of ACID4, indicated that the system made accurate predictions of lime requirements for maize (*Zea mays,* L.) and soybean (*Glycine max*, L.) but predictions for rice (*Oryza sativa,* L) and cassava (*Manihot esculenta*. L.) needed improvement (Dierolf et al., 1999).

Results from an exploratory, rule-based system FARMSYS (Colfer et al., 1989) illustrated that it was possible to merge multiple disciplines in a rule-based decision-aid. Ethnographic knowledge could be combined with knowledge of soil chemistry and management, when diagnosing and prescribing management when acid soil conditions were encountered. Local Minangkabau farmers preferred to grow rubber on their acid soils, which required no limestone applications and no tilling of the soil. Transmigrant Javanese and Sundanese farmers, on the other hand, would not hesitate to ameliorate their acid soils by applying the recommended limestone and tilling the soil for annual food crop production (Yost et al., 1992b).

Through repeated use of the decision-aid, users became familiar with typical requirements for particular crops, given usual levels of measured soil acidity, differences among soils and various crops. In fact, the users gained familiarity with the methodology, and learned certain aspects of the knowledge of managing acid soils. It is likely that some measure of the 'expert' knowledge was transferred to novice users through extensive use of the system. Perhaps the meta-level information was transferred to the decision-aid users as a result of using the system. It is clear, also that the detailed scientific knowledge was not transferred. Thus the mere use of the decision-aid does not replace the learning of the detailed algorithmic knowledge.

6.1.5 Observations

Further consideration of the factors affecting lime decisions indicated selection of lime materials could become impossibly complex. A linear programming model was developed that evaluated limestone cost, quality (fineness, neutralization capacity, as well as calcium and magnesium content), quantity available, and distance from the location for up to 5 limestone materials. These parameters were evaluated to provide a minimal cost choice of one or more of the limestone materials that met the specified soil pH and Ca and Mg targets in a spreadsheet decision-aid (Li et al., 1995).

While the main benefit of the decision-aid ACID4 was the use of the knowledge it contained, the process of organizing and recording the knowledge led to greater scrutiny of the knowledge and the identification of gaps and imprecision, which, in turn, led to improved subsequent research. This is illustrated in the evaluation of ADSS (a slight improvement over ACID4) (Dierolf et al., 1999). Thus, ironically, the preparing of the knowledge for dissemination, rather than detracting from the research process, actually improved and accelerated it. This meta-level analysis of the knowledge resulting from the crafting of the knowledge and representing it in the knowledge-base later proved to be extremely beneficial. This, in fact, may be a replication of the "patterns" and "larger framework" that experts seem to develop over time (Glaser and Chi, 1988)

6.1.6 Disadvantages

The ACID4 system provided a hands-on introduction to capture important knowledge and, for the Transmigrants of West Sumatra, critical knowledge about how to produce food on

these highly acid soils that differed so greatly from those of their experience. The system had several disadvantages including the following:

- The goal-driven, rule-based system proved rather unsuited to capture some of the information. In particular, social science information did not necessarily fit well in the rule-based knowledge representation system (Colfer et al., 1989).
- Many on-farm production limitations were due to multiple constraints occurring together. Acid soils in particular are characterized by multiple constraints. In addition to high acidity with toxic levels of aluminum and manganese. Levels of pH itself, calcium, magnesium, and phosphorus are to be expected to be insufficient and possibly yield limiting as well (Fox et al., 1985).
- A subsequent decision-aid was developed that attempted to address this problem (see section 6.4 NuMaSS, (Nutrient Management Decision Support System), later in this chapter).
- The system required a computer.
- This could be overcome by technicians and scientists running the software for the specific site or farm and communicating the results to the producer / grower.
- We later explore and propose a type of decision-aid that is completely graphic.
- Modification and updating of the software required rather expensive, proprietary software.
- One copy of the software could develop many systems (Le Istiqlal, 1986.)
- A small, free copy of the essential software was provided such that copies of the decision-aid could be copied and distributed inexpensively (run-time version).
- For subsequent decision-aids we used a procedure languages such as Pascal or declarative languages such as Prolog and hired programmers.
- Although the rules were given a numeric score of uncertainty, this uncertainty was combined in an inflexible way that often neither represented good practice nor the scientifically verifiable behavior.
- This effort led to subsequent improved representations of multiple sources of evidence (Bayesian cumulative probability) (Yost et al., 1999)---an implementation of evidence accumulation described in Pearl (1988).

Subsequent decision-aids included the cumulative probability to generate approximate confidence limits of numeric predictions of fertilizer needs using first order uncertainty analysis (Chen et al., 1997). This remains an area requiring more accurate representation of evidence accumulation as well as the appropriate handling of contradictory evidence. What are the most successful ways to carry out such calculations and accumulate evidence? It is likely that some of the methods recently used by IBM's Watson (Watson-Jeopardy, 201) would lead to better approaches than those described here. It also is not yet clear how successful experts make such estimates, if they do.

6.2 Propa (Papaya expert system)

That agricultural knowledge is highly interdisciplinary presents a challenge to the classical concept of an expert in a single discipline. When a grower or producer contacts the University with an issue they sometimes are referred to several experts before determining which expert is the right one for the specific problem. Confusion and failure to succeed in the diagnostic effort may occur. The goal of the Propa decision-aid was to explore this dynamic by attempting to construct a decision-aid that would identify and solve typical problems possibly requiring multiple disciplines (Itoga et al., 1990).

6.2.1 Goal
Develop an expert system comprised of multiple experts dealing with complex agricultural problems.

6.2.2 Objectives
Capture the knowledge of various scientists working with the papaya (*Carica papaya*) tropical fruit.

6.2.3 Implementing language
Prolog declarative language. Arity Prolog®.

6.2.4 Successes
The Propa decision-aid illustrated that it was possible for a group of experts from various disciplines to assess a case of a papaya problem and sort out which expert would be the primary expert to solve the problem. This was achieved through the use of a monitor and blackboard system that evaluated the interaction between the experts and the person with the papaya problem information. Each expert was assigned a dynamic relevancy factor which represented the success of their interaction with the papaya problem information. The disciplines brought together for the problem-solving session included experts in 1) Insect pests, 2) Nutrient management, 3) Disease identification, and 4) General management and good practice (Itoga et al., 1990).
Propa was able to show the user images of the various insects to assist and confirm their identification, which greatly assisted the insect expert's diagnosis and recommendation process.

6.2.5 Disadvantages
Test runs of the final system with users indicated that they were often overwhelmed with the number of technical questions that were asked of them by the group of experts. Many users were not prepared to answer dozens of questions about the detailed appearance of the plant and thus could not respond to the experts. When users could not respond to the expert's questions the experts were no longer able to proceed with a diagnosis.

6.3 PDSS (phosphorus decision support system)
The PDSS system development began in 1990 (Yost et al., 1992a).

6.3.1 Goal
Capture the knowledge, including both successful practice and the supporting scientific knowledge associated with the Diagnosis, Prediction, Economic Analysis, and Recommendations associated with managing nutrient phosphorus (P) in tropical food production systems.

6.3.2 Objectives
Capture a P management structure in a computer software decision-aid that would improve the management of the nutrient P.

6.3.3 Implementing language

Delphi® rapid application development software, Pascal language.

6.3.4 Successes

PDSS builds on the results of the structuring of the knowledge for the soil acidity decision-aid ACID4. As a result of the meta-analysis of the soil acidity decision-making process, we identified four components in the general process of nutrient management: 1)Diagnosis, 2) Prediction, 3) Economic Analysis and 4) Recommendation. These components served the basis for constructing PDSS and will now be discussed in succession.

6.3.5 Diagnosis

A diagnosis of a particular condition, in this case of a deficiency in soil and plant content of the nutrient phosphorus (P) is critical to bringing appropriate attention to the condition and, consequently, to its solution. A diagnosis in this sense can be observed when an expert is confronted by a problem and asks a few quick questions and rapidly determines the importance of further questioning or not. In this sense, the expert is exercising the "Best-First" search strategy discussed above. Such rapid assessments were observed when experienced scientists did field-visits, discussing with farmers the conditions of their crops. Often during such visits and discussions a suggestion resulted that led to corrective action. A diagnosis in this sense is our attempt to capture and implement an expert's best-first strategy of quickly assessing the seriousness of a situation and determining the best subsequent course of action. In another sense a diagnosis is a call to action. It is a decision about whether to act or not. This definition and use is important in terms of problem-solving and may be somewhat different than the classic "diagnosis" used in disease identification. The "diagnosis" we describe in this section is most effective if carried out by the person actually working with and intimately involved with managing the complex system (a crop-soil production system, in our case). A frequent heuristic or rule of thumb is that if a disease or condition is caught early then it is more likely to be successfully cured or remedied. Likewise, in complex systems of soil and crop management, a condition can often best be solved if it is detected early before subsequent, secondary complications, or in some cases irreversible damage, occurs. The analogy with human medicine is clear. For these reasons, it seems prudent for the grower, producer, or farmer to be informed and empowered with sufficient knowledge to detect the need for action. We also, upon further analysis, learned that there are other aspects of a good diagnosis that are important (Yost et al., 1999) (Table 1).

Diagnostic knowledge can be useful even if it is qualitative, highly observational, and even if a substantial amount of uncertainty is present. Highly uncertain information, when combined with other information with a similarly large amount of uncertainty, can, when taken together, begin to show a pattern that is typical of the disease, the condition or the state being detected. A good diagnosis could result from multiple pieces of information, none of which stands alone on its own, but when combined together, suggests a singular conclusion (i.e. all tending to indicate deficiency of a particular nutrient). We implemented this characteristic of being able to combine qualitative, quantitative, as well as uncertain information by using a Bayesian cumulative probability framework as indicated above in a chapter on Diagnosis (Yost et al., 1999). An example spreadsheet illustrating the calculations is shown in Appendix 2. The combining of multiple pieces of information thus often led to a

diagnosis when no individual piece of information was sufficient to provide a call for action. It was possible to include a consistency check, if mutually contradictory facts were observed. For example, if the probability of a nutrient deficiency for fact A was 0.9 that a P deficiency was likely (where 0 means certainty of no deficiency, 0.5 means complete ambivalence, and 1.0 means total certainty) and fact B had a probability of 0.2, then we have a situation of conflicting evidence. A rule was written to send a message to list in the output that a serious contradiction is occurring.

Table 1. Considerations in developing diagnostic questions. We suggest that the best diagnostic information/ tools/ questions are those that build on the common knowledge that on-site managers (e.g. farmers) have readily available together with simple measures, both qualitative and quantitative, of fundamental characteristics of the production system:

-The tool/question should be simple to use by lay persons.

-Results of the tool should be quick, such as the simple observation of a symptom or property in the field.

-Cost of the tool/question should be low or of no cost.

-The tool/question should be reliable as it should reliably indicate what action is to be taken.

Observations:

-Sometimes the result of the tool/question is that more expertise is required.

-Incomplete or imperfect data should not completely invalidate the diagnosis.

-The tool/question should take full advantage of the farmer, producer, or field observer's observation and knowledge.

-The tool/question may lead to improved, better diagnostic tools.

(Questions developed in a TPSS 650 Soil, Plant, Nutrient Interactions by students N. Osorio, X. Shuai, W. Widmore, R. Shirey. University of Hawai`i at Manoa)

We encountered two disadvantages of using the Bayesian accumulation of probability framework: 1) Much of our evidence and multiple observations or measurements were highly correlated or multicollinear. The multicollinearity contrasts with the assumed condition of independence in classic Bayesian evidence accumulation and thus the calculated cumulative conditional probabilities were in slight error depending on the degree of multicollinearity. 2) One could have strong evidence both for and against a condition as well as weak evidence for and against the condition, or even a complete lack of information, all of which would combine to a value of 0.5. As a result, strong, but conflicting, evidence is wholly discounted. One of our inadequate solutions to this situation was to monitor evidence and when evidence for and against a particular outcome differed substantially, a message was attached to the conclusion warning of the information conflict.

6.3.6 Predictions

The Prediction in PDSS is usually a numerical amount of a specified amendment needed to resolve the nutrient deficient condition identified in the Diagnostic section. There may be additional inferences based on the additional data usually required to complete the numerical prediction. There is a possibility that, upon analysis of the additional information, a prediction of no requirement may occur. The Prediction was developed using a combination of both scientific and local experiential knowledge. The preferred knowledge is that occurring when the best scientific methodology is gleaned from the literature and tested in the unique local soil, crop, weather, economic, and social conditions. To obtain and ensure such knowledge clearly requires intense work by the local scientists as well as the knowledge engineer (the person who organizes the knowledge and structures it into the knowledge representation format of the decision-aid software). In our case, scientists have included both international experts as well as local agricultural scientists who were in the process of or had completed field studies of the prediction methodology.

The choice of which knowledge and how much detail needed to be recorded and represented in order to minimize excessive detail and yet retain the essential knowledge was and seems to be a challenging one. This aspect has been lucidly discussed in Stirzaker et al. (2010). As Stirzaker et al. (2010) indicate, the typically detailed information resulting from research, needs to be smoothed and simplified to be most effectively used in a working system. Our experience has been identical and this aspect of building decision-aids seems to us to be one that requires close interaction and discussion with the intended users to best ascertain the appropriate level of detail. Thus it is clear that the direct transfer of algorithms and conclusions from a research effort is seldom possible without the requisite simplification described by Stirzaker et al. (2010). The intense and repeated contact between the developer and the client or user group has been essential in our experience. This type of intense interaction has come to be termed "extreme programming"(Beck, 1998; Wells, 2001). This programming style is based on frequent viewing and discussing of the software being developed with representative, intended users.

One of the requirements of the Prediction step that is necessary for the integration with the subsequent components is that there be a numeric prediction. This numerical value forms the basis of the benefit/cost analyses carried out in the subsequent Economic Analysis section. The Prediction equation of the PDSS decision-aid is thus an equation that began with the rather simple description given in (Yost et al., 1992a) shown in equation (2).

$$\begin{aligned} &\text{P requirement} = (\text{Soil P required} - \text{Soil P present}) \\ &/\text{Reactivity of the soil to added P} \end{aligned} \tag{2}$$

Where: P requirement is the kg/ha of fertilizer P needed to increase the soil P level ("Soil P present") to match the "Soil P required" and thus meet the specific crop's requirement for nutrient P. While equation (2) gives the basic structure of the P requirement prediction equation, there were updates to the equation which gradually increased in detail and complexity (eq (3)).

$$\begin{aligned} &\text{P requirement} = ((\text{PCL} - \text{Po}) / \text{PBC} + 0.8 * \text{PBC} * \text{Puptake} * 0.8 * 1/2) \\ &* \text{Placement factor} * \text{Application Depth} / 10 \end{aligned} \tag{3}$$

Where:

PCL = P critical level of the crop using a specific extractants ("Soil P required"of eq. 3)
Po = Initial, measured soil level of P using an specific extractant ("Soil P present" of eq. 3)
PBC = Phosphorus Buffer Coefficient using a specific extractant ("Reactivity of the soil to added P" of eq. 3)
Puptake = Yield of crop component removed*P content of the removed tissue (not present in eq. 3)
Application depth = Depth to which the fertilizer is incorporated (not present in eq. 3)
Placement factor = A factor that represents the relative efficiency of localized placement in reducing the P fertilizer requirement (not present in eq. 3)
The predictions developed in PDSS, as in ACID4, also included an expression of the associated uncertainty. In the ACID4 and FARMSYS modules the uncertainties were personal estimates of the reliability of the rules being exercised. In PDSS a different approach was used, that of error propagation (Burges and Lettenmaier, 1975). The error propagation calculation resulted in a very useful assessment of the equation's prediction. This was later expressed as the confidence limits of the prediction. An example of a prediction of P requirement was carried out on an experiment done at the Centro Internacional de Agricultura Tropical (CIAT) in Cali, Colombia and is illustrated in Figure 2. An interesting result of this prediction was that the actual precision of the fertilizer prediction was approximately +/- 50% of the requirement in most cases. This large error pointed out the typically large uncertainty in fertilizer predictions. One advantage of the first order uncertainty prediction was the ranking of sources of variability in the prediction equation. This enabled prioritizing research effort to better understand and make predictions (Chen et al., 1997).

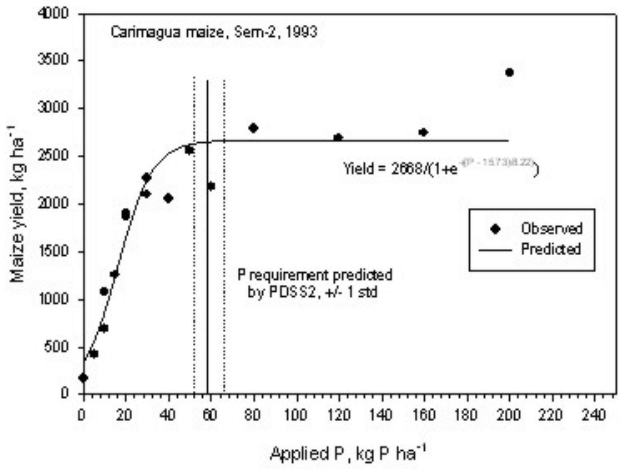


Fig. 2. Comparison of PDSS prediction with field estimates of the amount of fertilizer phosphorus needed to achieve maximum yield (CIAT, 1993)

6.3.7 Economic analysis

Economic analysis was the third component. This component clearly differed from the other portions of the decision making process requiring an economic calculation of profitability resulting from resolving the Diagnosed problem using the Prediction methodology. As

indicated above the construction of this component required quantitative output from the Prediction component in order to carry out the calculation of the benefit resulting from the solution of the diagnosed problem. This, for example, required a quantitative estimate of the amount by which the crop yield was increased as a result of supplying the necessary nutrient phosphorus. Understandably this required more than the usual soil science solution to increase extractable P. In PDSS, this required an estimate of crop growth response, which required estimating crop behavior at various levels of extractable soil P. This requirement meant that we had to incorporate plant response in addition to the simple chemical evaluation of extractable P. Thus we had to fill yet another knowledge gap in order to link the Prediction module with the Economic Analysis module. We found this "stretch" to ensure module communication helpful and broadening. As a result we gained an improved perspective of the decision-making process. The ultimate advantage was that we could conduct sensitivity analysis of the effects of change in extractable soil P on crop yield, profitability, and benefit/cost.

The adopted methodology for economic analysis was a simple partial budget analysis. This type of analysis permitted a quantitative calculation of benefit versus cost, giving some indication of economic advantage of the practice suggested in the Prediction step. The strength of the partial budget assessment was the minimal data requirement. The weakness, of course, was that the entire enterprise could be losing money but if the addition of fertilizer was resulting in yield increases considering fertilizer costs, the analysis would report a profit. Another weakness, from an anthropological point of view is that economic analyses capture only part of people's decisionmaking logic----issues like gender/ethnic division of labour and circular migration issues, symbolic meanings of particular crops, distaste for handling of fertilizers, food crop taste preferences, etc. are ignored...[of varying relevance, depending on local conditions]. In addition, the partial budget assumes no interactions among the fertilizer variables with other factors in the enterprise. Since the exploration of the consequences of various cost and benefit scenarios is often helpful for decision-support, a separate form was constructed to facilitate entry and calculation of benefit/cost given various price inputs.

6.3.8 Recommendation

The fourth and last component of the structure of nutrient management revealed as a result of the meta-level analysis of the decision-making process was the Recommendation. The Recommendation as identified in this analysis is the process and result of summarizing the entire decision-making process and includes the Diagnosis, Prediction, and Economic Analysis, and presents this information in a way that the decision-aids user can utilize. Understandably this varies with the needs, knowledge preferences and capabilities of the users. In the case of PDSS software a simple page is constructed that includes the specific segments of Diagnosis, Prediction, and Economic Analysis, and concludes with a list of the warnings (aspects of the consultation that could be seriously in error) or information notes that supplement the conclusions of the consultation.

The Recommendation, in the case of the SimCorn decision-aid (a decision-aid developed by scientists at Kasetsart University using the knowledge and algorithm implemented in PDSS) for the diagnosis, prediction, and economic analysis of fertilizer quantities for maize, was a book of recommendations that could be used by local extension officers to interpret soil test results and to communicate specific amounts of fertilizer blends for producers and growers in their region. In this case, the extension officers provided the information verbally rather

than distributing leaflets and tables of fertilizer recommendations (Attanandana et al., 2007; Attanandana, 2004; Attanandana and Yost, 2003; Attanandana et al., 2007).
Preparing the decision-aid knowledge for the Recommendation thus requires close contact and familiarity with the clients, or with the agents who will be the direct users of the software. As discussed in Attanandana et al. (2008), the results of the decision-aid consultation should be prepared in a form that enables and empowers the producer/farmer who will be using the results. The preparation of the Recommendation thus completes the process of close contact with the eventual user of the decision-aid results that we consider essential for the crafting and construction of the decision-aid as well as its application (Attanandana et al., 2007; Attanandana et al., 2006).

6.3.9 Reaction to the PDSS decision-aid among differing collaborators

The PDSS system and the knowledge contained therein was found useful in various ways by our collaborators. For example, our Thai colleagues sought to include PDSS for the P algorithm contained therein. They incorporated the logic and equation into their own systems, SimCorn and SimRice (Attanandana et al., 2006). Our colleagues in the Philippines, however, preferred to receive the PDSS algorithms in the form of the more integrated NuMaSS software, to be discussed subsequently, which combined the nitrogen, phosphorus, and soil acidity components (Osmond et al., 2000).
The use of the PDSS algorithms in our collaborators' software SimCorn (Attanandana et al., 2006) reduced the recommended application of phosphorus by roughly 50% (Attanandana, 2003, personal communication) reducing the requirement for foreign exchange to purchase fertilizer P, and limiting the accumulation of environmentally harmful levels of nutrient P.

6.3.10 Expansion and extension of the PDSS decision-aid

The PDSS decision-aid, first released in 2003, proved to be a decision-aid in development. The development of PDSS, similar to the development of ACID4, opened up new possibilities and suggested several additions and generated multiple research activities. The areas where additional knowledge was prioritized included the addition of a potassium module especially for work in Thailand. Also in Thailand and in West Africa, we needed to help identify rock phosphate-favorable conditions as well as the amounts that should be applied to alleviate P deficient conditions. And, lastly, we needed to diagnose and predict nutrient requirements in perennial cropping systems such as trees, which was clear from the intial work with decision-aid ACID4 in Sumatra, Indonesia.

6.3.11 Improving predictions of the PDSS

As a result of calculating the error in the prediction (Chen et al., 1997), which gave confidence limits on the decision-aid prediction, we also obtained the relative ranking of error in each of the input variables. This information was then used to identify the greatest source of error in the prediction. This led to the identification of follow-up research designed to reduce error and uncertainty in the predictions. Follow-up work was carried out, for example, to better estimate and predict the buffer coefficient for phosphorus in various project sites (George et al., 2000).

6.3.12 Potassium module

Another substantial gap in the nutrient management of crops for food and fuel in the Tropics included the need to assess the potassium (K) status of highly weathered soils. We

expect deficiencies in potassium to occur and indeed our experience has been exactly that in Thailand (Attanandana and Yost, 2003). With assistance from our collaborating institution and local research support, a study of methods for diagnosing and predicting potassium requirements was completed. Based on this result and when integrated with other preliminary research, a tentative potassium prediction model was proposed (Yost and Attanandana, 2006), eq. 4.

$$\begin{aligned} &\text{K requirement}\left(\text{kg K ha}^{-1}\right) = \left(\text{Kcritical} - \text{Ksoil}\right) / \text{BCK x B.D. x} \\ &\left(\text{Application depth} / 10\right)\text{x Placement factor} + \\ &\left(\text{Biomass removed x K content in the biomass}\right) \end{aligned} \tag{4}$$

Where
K requirement = the amount of fertilizer K that is needed to restore the soil K supply such that crop yields were maximum
Kcritical = The level of soil K needed to ensure that maximum growth and productivity occurred
Ksoil = The measured level of soil K
BCK = The soil buffer coefficient, i.e. the reactivity of the soil to added K, using the same extractant as Ksoil
B.D. = Soil bulk density (specific gravity), i.e. the weight of soil per unit volume
Application depth = The intended depth of incorporation of the fertilizer K in cm
Placement factor = A fraction that represents the relative benefit from application to a fraction of the soil volume at the specified depth to be fertilized
Biomass removed = The amount of crop bioproduct that is expected to be regularly removed from the field
K content of the biomass = The K content of the portions of the crop that will be removed from the field
Subsequent comparisons of yield and profit from farmer practice as compared with decision-aid recommendations indicated yield increases where K was applied according to predictions and increases in profit (Attanandana et al., 2008). Further and more detailed studies indicated that new methods for K diagnosis should be considered (Nilawonk et al., 2008).

6.3.13 Rock phosphate module

Another substantial gap in the nutrient management of crops for food and fuel in the Tropics included the need to consider locally available sources of nutrient phosphorus. This was an issue both in Thailand and in Mali, West Africa. A systematic analysis of the issues and factors that control rock phosphate effectiveness was carried out and the results were organized into a decision-tree and logical sequence (Sidibé et al., 2004). This author proposed a comprehensive approach to determining whether and how much of a specified rock phosphate material should be applied to restore crop productivity. The result was an algorithm that successfully predicted rock phosphate applications in acid sulfate soils of Thailand (Yampracha et al., 2005).

6.3.14 Perennial crops module

Initially, the project anthropologist observed that local communities had long preferred perennial crops for a variety of reasons later to become apparent (Colfer, 1991; Colfer et al., 1989).

Subsequently, it was apparent that in some high rainfall tropical environments the repeated clearing of land for food crops resulted in exposing bare soil to the intense rainfall. This could result in damage to the soil status either by leaching and loss of soluble nutrients on one hand (Dierolf et al., 1997), loss of enriched, surface soil through soil erosion, or both. In certain environments a more conservation-effective agro-ecosystem may be a perennial production system that provides regular food production but where the soil surface remains covered and protected from the typically highly erosive rainfall of humid tropical environments, such as those of tropical Indonesia.

A meta-level analysis of nutrient management structure and options in perennial cropping systems suggested that there also was a discernable and distinctive structure in perennial cropping systems and that the structure included the following: 1) A nursery phase, in which the seeds of the perennial plant were germinated and young plants begun, 2) An establishment phase in which the small seedlings were outplanted into the land that would become a forest, 3) A period of fast growth, and 4) A mature phase, in which production continued for many years (Figure 3). A review of the literature was assembled considering the perennial producer peach palm (*Bactris gasipaes*, L.) (Deenik et al., 2000).

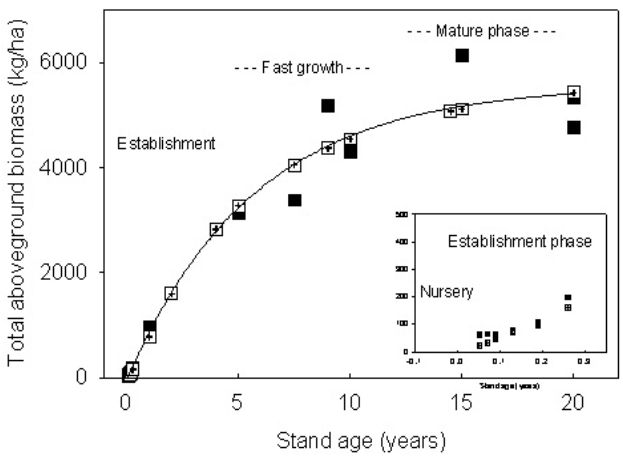


Fig. 3. A graphic depiction of four stages of particular importance in the management of nutrients in perennial crops (Deenik et al., 2000).

As in the case of other meta-level analyses of the agricultural systems, numerous gaps were observed in the available nutrient management in perennial crops when viewed from this perspective. As a result several studies were undertaken including collaborators, especially of the University of Costa Rica and the EMBRAPA/Amazon center in Brazil (Ares et al., 2002a; Ares et al., 2002b; Ares et al., 2003; Yost and Ares, 2007).

6.3.15 Summary

The meta-level assessment of patterns and structure, found so helpful in the development of ACID4, was quite helpful in the course of developing PDSS. The meta-level analysis of the knowledge base resulted in numerous improvements in the concepts and content of nutrient management in the body of knowledge (knowledge-base) surrounding the nutrient phosphorus in tropical ecosystems. The identified structure (Diagnosis, Prediction,

Economic Analysis, and Recommendation) identifies major components in the decision-making process and helps the user step through the process.
Developing a tool that predicts actual quantities of fertilizer needed given specified data illustrates that multiple disciplines can contribute to solutions in a systematic, synergistic way provided a common language is used. Is it true that computer-based knowledge systems can better accommodate and use multiple disciplines than humans?
A structure was proposed that enabled a monitor to guide and organize a multidisciplinary search to solve an unidentified problem in papaya problem diagnosis.
The calculation of propagated error in the prediction of fertilizer quantity at approximately 50% illustrates that despite attempts to achieve highly precise estimates in agronomic research, the results need improvement to better support economic and environmental objectives. The calculation of propagated error identified the variables most contributing to error. These variables became objectives for research to improve prediction accuracy and precision.
Conducting research to produce a knowledge module provided a clear, stimulating objective and resulted in high quality, problem-driven, and yet rewarding research.
The meta-analysis of the decision-making process and surveys of the users resulted in the identification of knowledge gaps that, in turn, resulted in numerous educational opportunities for young scientists. They conducted their research knowing that the results of their work would fill a gap in a working knowledge-base. The following scientists completed dissertations and Ph.D. degrees during this work (Gill, 1988; Evensen 1989; Agus 1997); Deenik, 2000; Diarra et al., 2004; Dierolf et al., 1999; Nilawonk et al., 2008; Sipaseuth et al., 2007; Yampracha et al., 2005).

6.4 NuMaSS (nutrient management support system)

The NuMaSS Project was designed to join the individually developed decision-aids ADSS and PDSS, with a new system to be adapted from a nitrogen decision-aid (Osmond et al., 2000).

6.4.1 Goal

The NuMaSS Project was developed to integrate and disseminate decision-aid tools that diagnose soil nutrient constraints and select appropriate management practices for location-specific conditions. The strategy was to develop globally applicable, largely computer-assisted, integrated decision-aids that could both diagnose and prescribe appropriate solutions to soil nutrient constraints.

6.4.2 Objectives

The Project had three objectives 1) Improve the diagnosis and recommendations for soil acidity and nutrient problems. 2) Develop an integrated computerized knowledge-base, and 3) Develop auxiliary tools resulting from the integrated knowledge-base to assist producers to diagnose and solve soil acidity and nutrient constraints.

6.4.3 Implementing language

The rapid application development package Delphi®, which was a Pascal-based development environment.

6.4.4 Successes

The integrated software was comprised of existing modules of PDSS and ADSS, both modified to merge into the integrated framework called NuMaSS (Nutrient Management Support System). A nitrogen module was added by North Carolina State University based on the Stanford (1973) approach of determining a balance of what the crop needed relative to the soil content with amendments and fertilizers added. The actual implementation, while based on the Stanford approach, grew out of a dissertation of one of the project members (Osmond, 1991). The NuMaSS software was disseminated in countries in West Africa, Central and South America and in S.E. Asia (Thailand and the Philippines).

One of the notable initial successes of NuMaSS was that by following the diagnosis of soil acidity conditions and following with the proper soil amendments, desired crops such as mung bean (*Vigna radiata,* L.) could be grown where the crop had died previously due to the high soil acidity (Aragon et al., 2002). The decision-aid also indicated that other crops could be grown and would require substantially less expensive limestone than did mung bean. The initial success continued and gradually attitudes and awareness towards soil acidity changed and producers became aware of the importance and limits in productivity it caused. Several assessments of farmer attitude indicated substantial change in awareness and prioritization of soil acidity (Aragon et al., 2002). An impact analysis at the conclusion of the project reported that during the next 40 years the project results were conservatively estimated to return about 45 million $US in benefits to the producers (Walker et al., 2009).

Some other spectacular results occurred on the island of Negros Occidental where farmers and producers had not been applying limestone and were basically unaware of soil acidity. In this province maize yields of over 7 metric tons were obtained with the addition of nutrients according to NuMaSS predictions. This contrasted to yields of maize of 1 to 2 tons without the addition of nutrients or limestone (D. Cidro, 2006 personal communication).

The impact of the introduction of NuMaSS and introducing specific management of the acid, uplands soils seems on track to expand and extend well beyond the province of Isabela where the Walker et al. (2009) study took place. Other provinces of Northeastern Luzon have begun instituting province-wide programs of applying limestone. This contrasts to the total lack of commercial agricultural limestone in the regional city of Ilagan when the project begun. It was not clear that the rapid and extensive adoption of the liming technology was due to NuMaSS, but it seems likely that the dissemination was enhanced by the presence of the decision-aid.

6.4.5 Disadvantages

While it is clear that the NuMaSS software assisted and improved food crop yields in the Philippines, it was also clear that there were problems with the nitrogen component, especially in Thailand and Laos. A dissertation study was carried out comparing N recommendations from two decision-aids (NuMaSS and DSSAT (Jones, 1998)). The results indicated that neither software adequately estimated the minimum amounts of fertilizer N that should be applied. A dissertation study indicated that there was substantial residual nitrate that should be measured and which reduced the fertilizer nitrogen requirement (Sipaseuth et al., 2007).

Unfortunately, the results of the multiple expert work described in the Propa system (see section 6.2) had not yet become available and separate, non-interacting systems for nitrogen,

phosphorus, and acidity were constructed. In addition time did not permit the full integration of the potassium, rock phosphate, and perennial crop modules that had been developed for PDSS, to be integrated into the NuMaSS system.

6.5 NuMaSS-PDA (nutrient management support system for personal digital assistants)

Scientists at Kasetsart University, Bangkok, Thailand were among the first to attempt to diagnose and make fertilizer recommendations at the field-level using decision-aids (Attanandana et al., 1999). These scientists adapted simulation model output for use by local growers / producers in their efforts to apply site-specific nutrient management on their land (Attanandana and Yost, 2003). This approach was an attempt to adapt the concepts of Precision Agriculture, which broadly seeks to apply the right amount of the right nutrient at the right place at the right time, to the small farms of the Tropics (Attanandana et al., 2007). This included invention and use of the soil test kit, identification of the soils in the field using simple field observations (Boonsompopphan et al., 2009) and simplification of complicated simulation models (Attanandana et al., 2006; Attanandana et al., 1999) so that fertilizer recommendations could be made in the field. These efforts led to the assembly of the SimCorn, SimRice, and subsequently to the SimCane software (Attanandana et al., 2006). The NuMaSS-PDA was an attempt to harness the new capability of handheld computers and smartphones to provide the decision-support for the process.

6.5.1 Goal

Re-structure and re-organize a subset of the knowledge in the NuMaSS software for delivery on a hand-held computer so that one could go to a producer's field, sample and analyze the soil, identify the pertinent soil series, and conduct the diagnosis, prediction, economic analysis, and prepare a recommendation on site.

6.5.2 Objectives

Adapt essential parts of the NuMaSS decision-aid for delivery on a handheld device. Add the potassium decision-aid module to that of the nitrogen and phosphorus, thus providing direct support for the typical N, P, K fertilizer decision-making. Develop a simple method so that interaction and use of the decision-aid would be possible in multiple languages with simple addition of a simple dictionary for each language.

6.5.3 Implementing language

Superwaba®, a Java-based language. Palm OS, Windows Mobile OS.

6.5.4 Successes

A multilingual interface was developed that would permit interaction with users from a large number of languages. The nitrogen, phosphorus, potassium and liming modules were joined and executed properly in the SuperWaba environment. The initial languages implemented included English, French, Portuguese (Li et al., 2006). Subsequently, Tagalog and Tetun were added. Development concluded as the project drew to a close.

6.6 Visual decision-aid

In numerous regions of the Tropics access to a computer, an agricultural officer, or highly knowledgeable producers / growers is difficult or impossible. In other cases, even access to

written literature does not provide access to the knowledge written there. Nevertheless, producers and growers are intensely aware of, and clearly survive on visual information. We began an attempt to explore the possibility of stimulating awareness and transfer of information, or learning by a completely visual approach. This was a completely visual guide to the installation of a water conservation technology called "Amenagement en courbes de niveau" (ACN) in West Africa. This water conservation technique was developed by CIRAD (Gigou, 2006) and later characterized (Kablan et al., 2008). Adoption of the technology was slow, in part, due to the requirement for expert delineation of the contours in the field. Demand for the technology far out-stripped the availability of the local scientific staff. Professional staff are required to survey the hydrological issues in the field and devise strategies to handle the issues and locate the contours. The visual aid, Visual ACN, illustrated in Appendix 3, is proposed to illustrate, inform, and instruct in the installation and maintenance of the technology including the illiterate, which often may be women farmers, for example. The decision-aid in this case was not a computer software; rather it was a simple guide based on a sequence of two figures or drawings per page illustrating the condition and the solution on each page. We have conducted test distributions of the guide, which have stimulated substantial interest among local producers and growers (Figure 4).

Fig. 4. Long time, successful ACN (Amenagement en Courbes de Niveau) user, M. Zan Diarra, Siguidolo, Mali, West Africa examines the Visual ACN decision-aid with interest (photo: R.Kablan). How to assess what he might be remembering or learning from the visual information?

6.7 Decision-aids to assist learning / education

Use of the decision-aids was found to stimulate and enable learning in several ways.

- Operating the decision-aid by entering the requisite data. Carrying out the calculations and recording or transmitting the results helped users gain familiarity with the software and elements of the decision-making process. Merely operating the software did not result in the users gaining the knowledge embedded in the software, but to the curious it led to questions and sometimes sparked curiosity.
- Observant users gained some measure of the types of information needed and could test the system integrity by checking test cases. A feature of the early expert system shells was back-tracking that demonstrated the logic that led to a specific conclusion. With later decision-aids this feature was not present so supplementary information surrounding the consultation had to be selected and specifically added at the Recommendation stage.
- The rigor imposed by the decision-aid on the user, i.e. having to provide an answer in every run or consultation, reinforced and reminded the users of the need to have information from a specified group of interdisciplinary knowledge bases. The input from farmers as local experts, both improved the system and provided positive feedback to them.
- For some users the rigor of having to answer, in a precise way, was difficult and resulted in loss of interest and failure to complete the consultation.
- Use of a decision-aid can illustrate to users the need for and value of interdisciplinarity.
- Exposure to the way knowledge can be organized provides an exposure to a meta-level appreciation of the problem-solving techniques, which can support learning.
- The structure of the decision-making process illustrates one type of problem-solving that users can adopt or modify for themselves in the future as needed.

6.8 Summary

This chapter illustrates that complex agricultural knowledge could be captured and implemented so that numerical predictions and informed recommendations could be produced. In the example given, soil and crop management technology and knowledge on acid soils was captured from successful practice in the Southeast US, Central and South America and implemented in analogous soils in Indonesia.

Use of the decision-aid by users in a new location where soil acidity or phosphorus was limiting helped users identify data needs for solving such problems. For example, improved management of acid soils acid soils in the uplands of the Philippines was stimulated by the introduction of the decision-aid NuMaSS. The improved management included introducing the practice of liming. As a result producers and growers in the region were expected to benefit over 45 million $US, according to the impact analysis conducted in 2007.

The knowledge engineering process led to a meta-analysis of the process, i.e. some thought about how to best solve nutrient management problems. The result of the meta-analysis was the identification of a structure in nutrient management decision-making. A structure of Diagnosis, Prediction, Economic Analysis, Recommendation was proposed. This structure guided the formulation of PDSS, NuMaSS, and NuMaSS-PDA.

6.9 Challenges

Improved knowledge structures are needed that better match the nature of the knowledge.

Improved methods of knowledge representation are also needed such that the knowledge does not die when a software falls out of use or is discontinued. The use of pseudo-code or alternative knowledge capture may be an useful in this regard.
The knowledge engineering process is expensive and time consuming. The knowledge capture process needs further streamlining so that non-experts can record and exercise their knowledge.
Better software tools are needed to enable decision-aids to serve the learning / teaching function.

6.10 Acknowledgments

Numerous persons in addition to the authors have made development of the decision-aids discussed in this chapter possible. We list all of our names, knowledge areas, and institutions here.
Authors:
Russell Yost, Soil Scientist, University of Hawai`i at Manoa, Department of Tropical Plant and Soil Science, Honolulu, Hawai`i
Tasnee Attanandana, Soil Scientist (retired), Kasetsart University, Department of Soil Science, Bangkok, Thailand
Carol J. Pierce Colfer, Anthropologist and Senior Research Associate, Center for International Forestry Research (Bogor, Indonesia) and Visiting Fellow, Cornell Institute for Food, Agriculture, and Development, Ithaca, New York.
Stephen Itoga, Professor of Information Technology (retired), University of Hawai`i at Manoa, Department of Information and Computer Science, Honolulu, Hawai`i.
Other Contributors.
Tsuji, Gordon; Program Coordinator and Financial Officer, University of Hawai`i at Manoa, Departments of Agronomy and Soil Science and subsequently of the Department of Tropical Plant and Soil Sciences.
Zhi-Cheng Li, Senior Program Analyst, Stanford Knowledge Project, Stanford, California
Madonna Casimero, Former Project Coordinator, SM-CRSP, The Philippines, IRRI, Los Banos, The Philippines
James B. Friday, Extension Forester, University of Hawai`i, Department of Natural Resource Management, Hilo, Hawai`i
Kouang Doungsila, Senior Scientist, National Agriculture and Forestry Research Institute, Vientiane, Laos.
Nivong Sipaseuth, Research Scientist, National Agriculture and Forestry Research Institute, Vientiane, Laos.
Mamadou Doumbia, Director LaboSEP (Laboratory of Soil, Water and Plants), Institut d'Economie Rurale, Bamako, Mali
Aminata Badiane, Natural Resource Manager, USAID/Senegal, Dakar, Senegal
Modou Sene (retired), Scientiste du Sol, CEERAS , Thies, Senegal
Aminata Sidibe-Diarra, Natural Resource Manager, USAID/Bamako, Mali
Ricardo Maria, Soil Scientist, Instituto de Investigações Agronómicas de Moçambique, Maputo, Moçambique
Candida Cuembelo, Soil Scientist (retired), Instituto de Investigações Agronómicas de Moçambique, Maputo, Moçambique
Felicidade Jorge, Professora de Ciencia do Solo, Universidade de Eduardo Mondlane, Maputo, Moçambique
Istiqlal Amien, Senior Soil Scientist, Center for Soil, Agroclimate Research and Development, Bogor, Indonesia

Fahmuddin Agus, Doctoral student, N. Carolina State University, and Centre for Soil and Agroclimate Research and Development, Bogor, Indonesia. Tom Dierolf, Senior Scientist, formerly of the Heifer International and International Consultant
Dan Gill, Doctoral student, N. Carolina State University.
Michael Wade, Senior Scientist, N. Carolina State University
Richard Kablan, Senior Research Associate, Department of Tropical Plant and Soil Science, University of Hawai`i at Manoa, Honolulu, Hawai`i
This study was was made possible by the generous support of the American people through the United States Agency for Development, International Development, under the terms of Award No. LAG-G-00-97-00002-00. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development.

6.11 Abstract

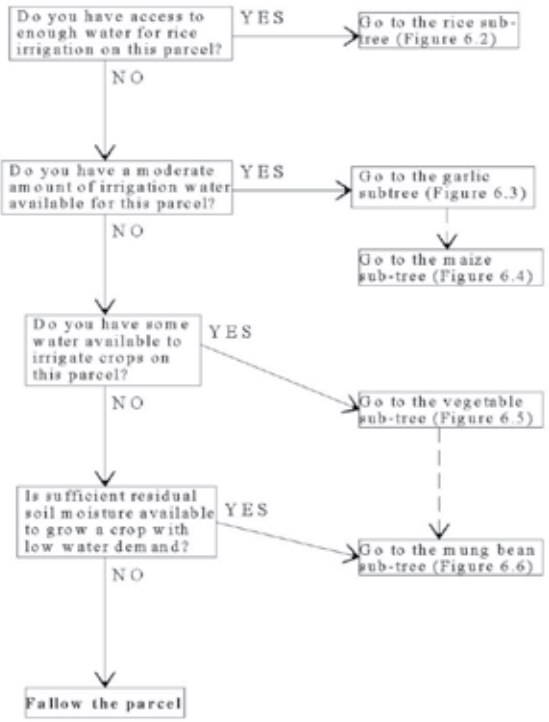
The application of Information Technology to agricultural knowledge in the form of decision-support systems and decision-aids has already been successful and offers great promise of more success in the future. This chapter will discuss the complex characteristics of agricultural information and the challenges it presents to society and to Information Technology (IT) to capture, organize, and disseminate this knowledge. Successes resulting from the application of IT to agricultural knowledge are discussed. The challenging characteristics of agricultural knowledge are that it is 1) Highly experiential and highly situational, 2) Characterized by outcomes and results that are risky, uncertain, with many events minimally predictable, 3) Interdisciplinary in nature, ranging from social sciences to biology, chemistry, physics, and mathematics. That this knowledge has been an extraordinary challenge for IT to capture is well-known and it is not surprising that new methods, structures, and systems continue to be required to capture the information.

Agricultural knowledge remains in a much earlier more descriptive stage of development, where personal observation and personal experience play a primary role in understanding agriculture and attempts to control or manage it. It is the thesis of this chapter that information technology can play a role in moving agriculture along to a more advanced stage of development by recognizing consistent trends, patterns, rules of the trade, rules of thumb and building upon such knowledge. Meta-analysis of the state of agricultural knowledge should be encouraged as being helpful to the process. The understanding of the causes of risk and uncertainty is relatively recent as are the benefits from systems studies conducted at a relatively high level of abstraction. Some examples include the identification of "tipping points" and their recognition in fragile ecosystems. Other aspects of the mathematics of catastrophe theory seem to offer benefits of perspective and overview of complex systems such as agriculture. The highly interdisciplinary nature of the knowledge, while involving virtually the full range of human knowledge systems, goes well beyond the scope of traditional biological, physical, and chemical disciplines. Of particular note is the importance of social sciences in the understanding and in attempts to control and manage real agricultural systems. As indicated, the need for decision-support systems to consider the human element in observing, capturing, and delivering management expertise is well known and several examples have been given. Preliminary exploratory decision-aids are discussed and the results of the development effort are chronicled in this chapter. One example depicts the capture for knowledge associated with the management of acid soils, developed in the SE US, Central America, and South America, and enable its transfer and

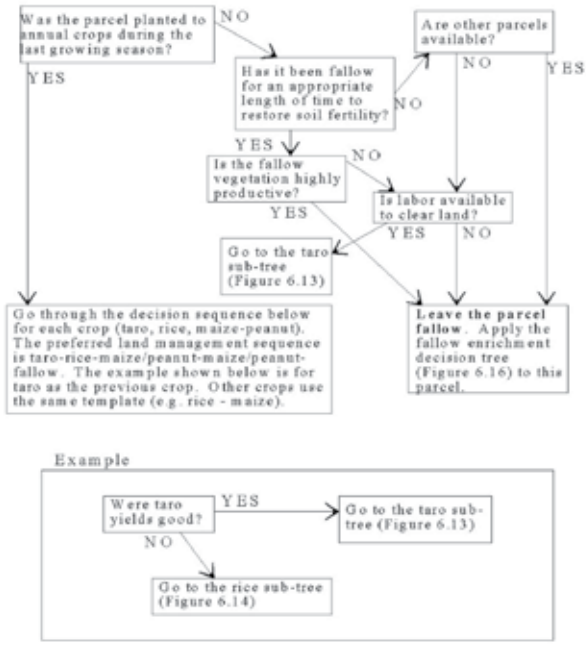
adaptation by farmers of upland cropping systems in the province of Isabela, The Philippines. There acid soils abound but techniques of acid soil management were not to be found. The results of the capture of both the experiential and scientific knowledge led to the identification of patterns and structures in the knowledge. Four components of nutrient management decision-making were identified and proved to be helpful in understanding, predicting, and controlling the management of nutrients in tropical soils. These included the Diagnosis, Prediction, Economic Analysis, and Recommendation components of nutrient management decision-making. Numerous gaps in both knowledge representation, knowledge organization, and in the use of decision-aids to support and transform teaching and learning remain. Nonetheless, it seems the potential of information technology is yet to be understood and certainly not realized.

7. Appendix 1

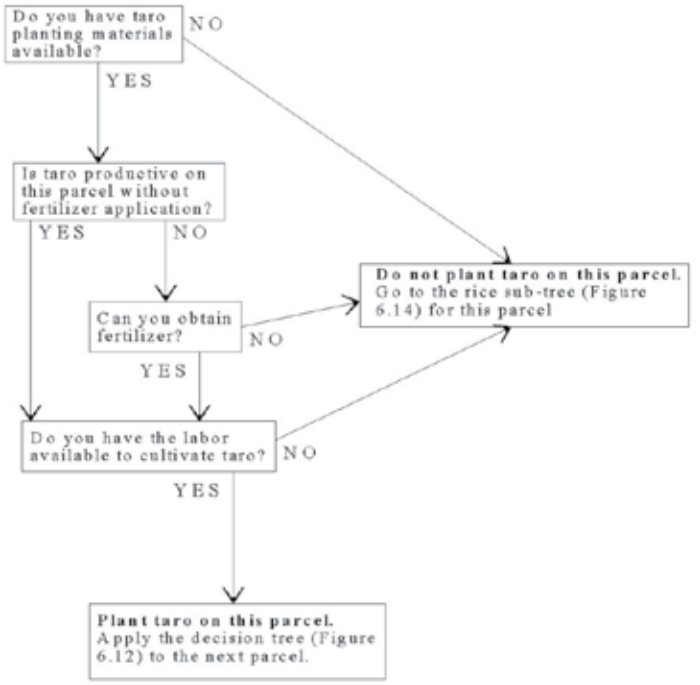
Example synthesis and inference of guiding decision-trees of agricultural information (Robotham, 1998).



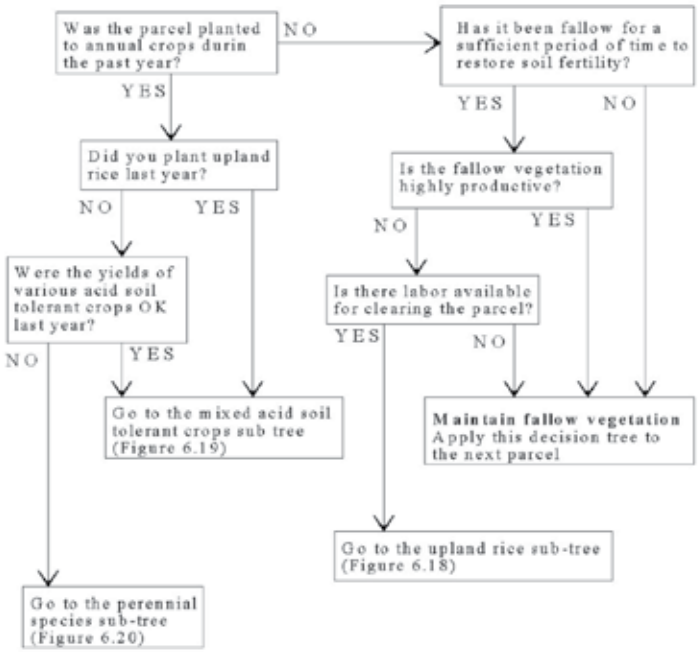
Appendix Figure 1(Robotham, 1998, Figure 6.1).



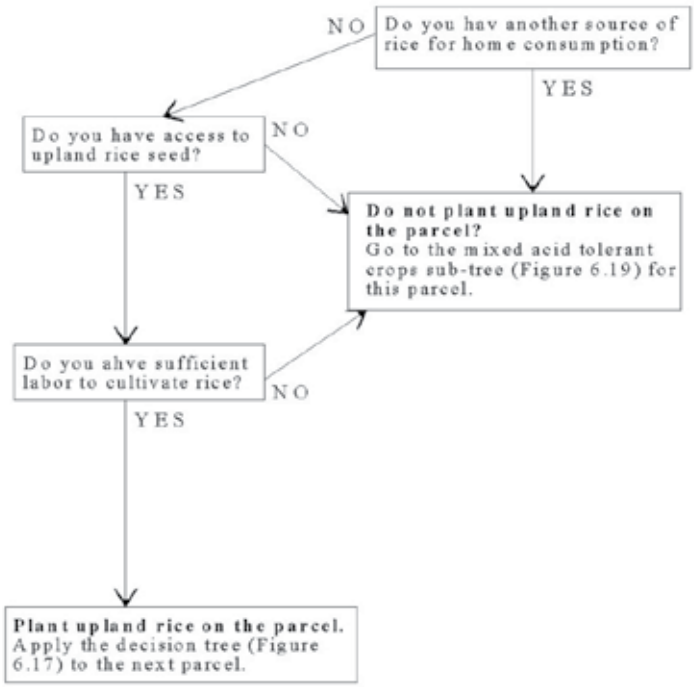
Appendix Figure 2 (Robotham, 1998, Figure 6.12).



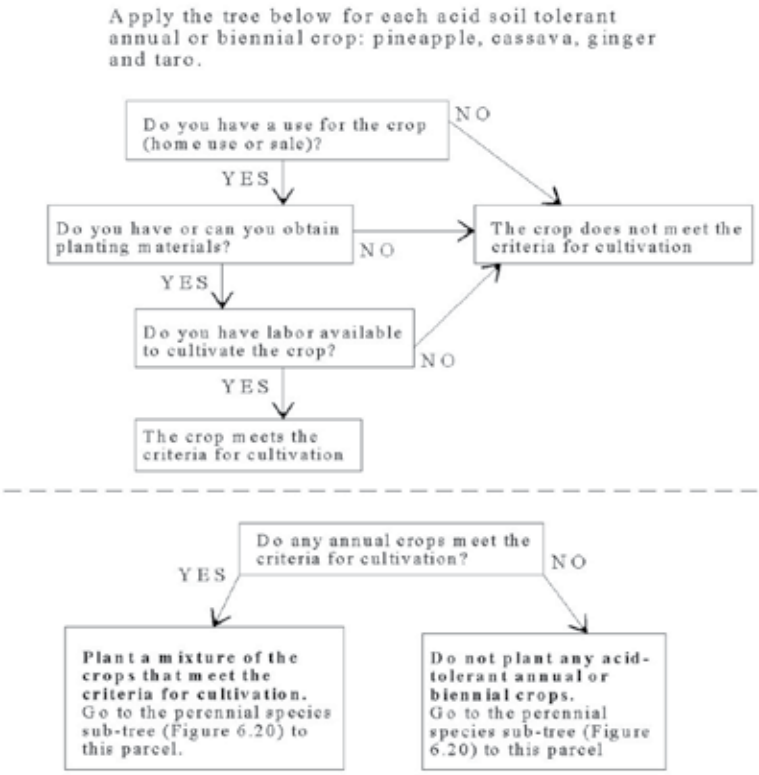
Appendix Figure 3 (Robotham, 1998, Figure 6.13).



Appendix Figure 4 (Robotham, 1998, Figure 6.17).



Appendix Figure 5 (Robotham, 1998, Figure 6.18).



Appendix Figure 6 (Robotham, 1998, Figure 6.19).

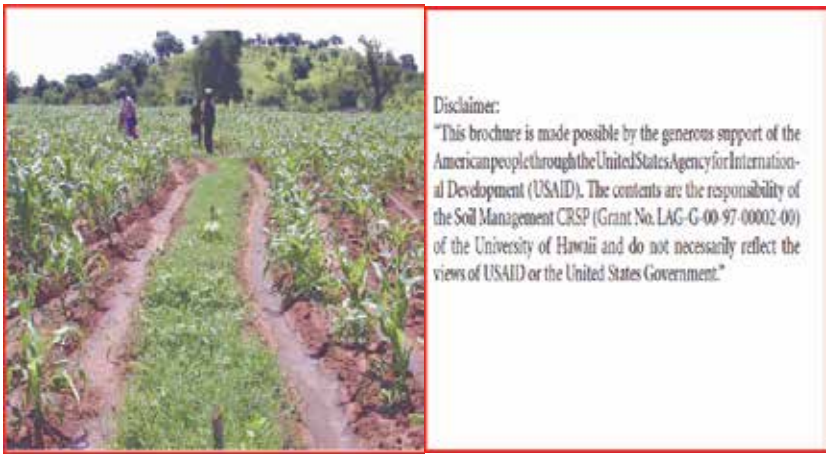
8. Appendix 2

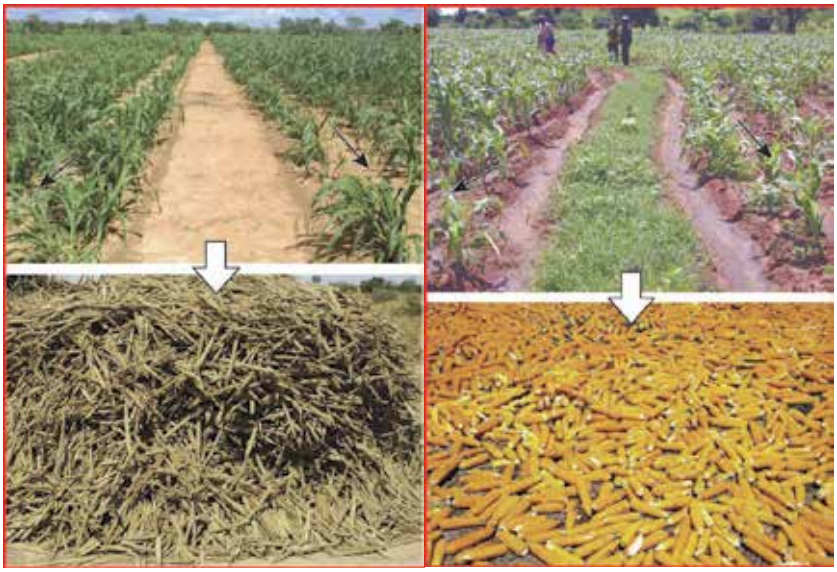
Example table illustrating the calculation of cumulative probability. Initial probability of P deficiency was considered 0.5, which on a scale of 0 to 1 indicates no information for or against P being deficient in the examined soil. The corresponding odds of a deficiency then is 50/50 or equal to 1. As various factors are considered, each with a probability of P deficiency (P(def)), the evidence accumulates until a final cumulative probability of 0.83, which indicates a high probability of P being deficient in the measured soil soil (Yost et al 1999; Pearl 1988).

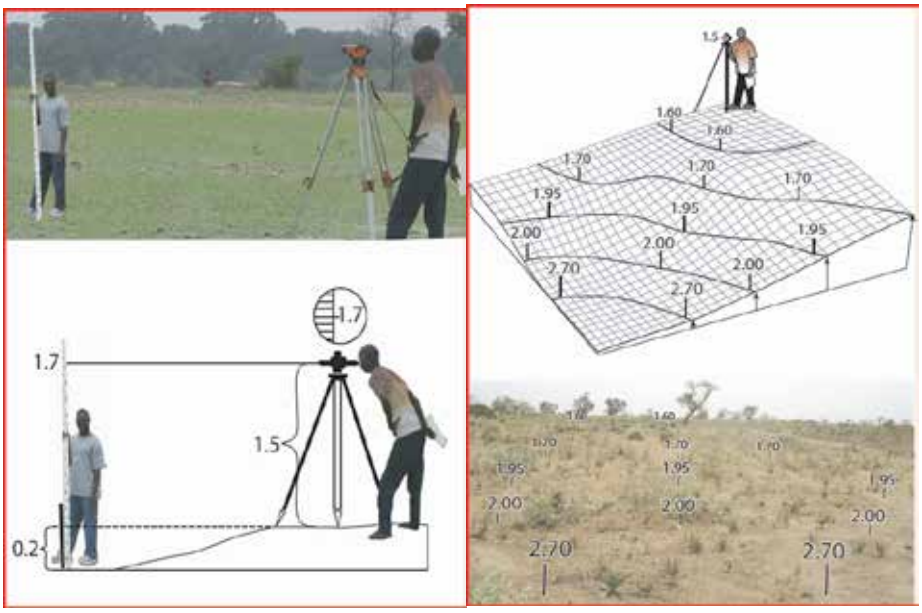
Example Bayesian calculations of cumulative probability					
Initial P deficiency =	0.5	Odds (deficiency):1			
				Odds	Cum.Prob
Factor	Value	P(def)	L(def)	(def)	(P deficiency)
Agric. Region	Mali	0.6	1.50	1.50	0.60
Soil Order	none	0.5	1.00	1.50	0.60
Prev. Crop	none	0.5	1.00	1.50	0.60
Indicator Plant	Striga	0.7	2.33	3.50	0.78
Def. Symp.	Purple	0.7	2.33	8.17	0.89
Plant Anal.	0.3	0.2	0.25	2.04	0.67
Soil Anal.	< 0.5	0.7	2.33	4.76	0.83
Total Cum Probability:					0.83

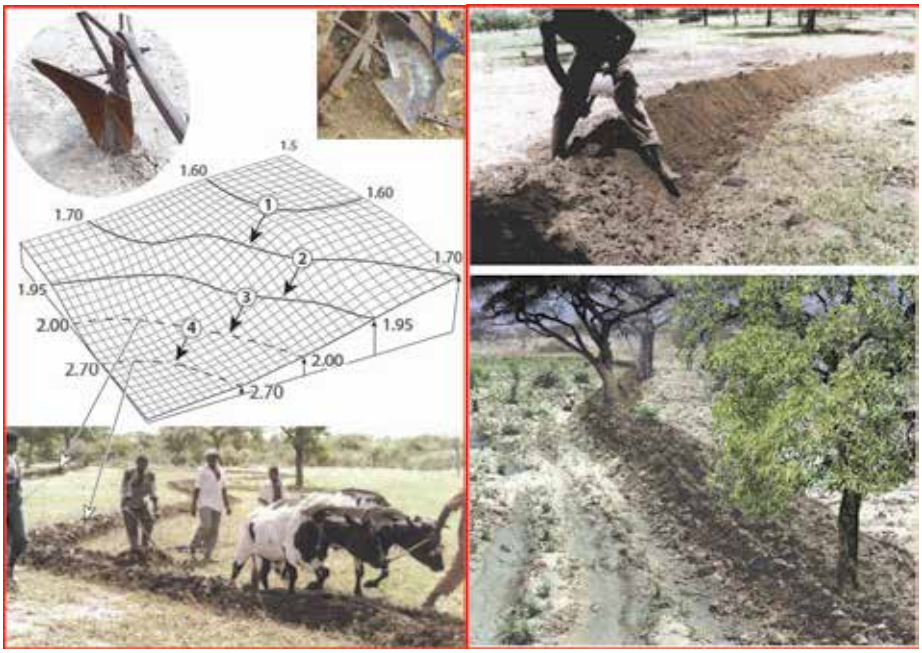
9. Appendix. 3

Visual ACN






1.7
1.7
1.5
0.2
1.5
1.60
1.60
1.70
1.70
1.70
1.95
1.95
1.95
2.00
2.00
2.00
2.70
2.70
2.70
2.70


1.5
1.60
1.60
1.70
1.70
1.95
1.95
2.00
2.00
2.70
2.70
1
2
3
4


1
2
3
4
5

7. Reference

Aragon, M.L., J.B. Lasquite, T. George, T.M. Corton, and R.S. Yost. 2002. Development of Decision Aids for Integrated Nutrient Management Support System (NuMaSS) for Rice-based Systems in Acid Uplands. M Casimero, M Aragon, J Lasquite, and R Yost, eds, Regional Workshop in Decision Aids for Integrated Nutrient Management Support System (NuMaSS) for Rice-Based Systems in Acid Upland Areas (SM-CRSP).

Ares, A., J. Boniche, E. Molina, and R. Yost. 2002a. Bactris gasipaes agroecosystems for heart-of-palm production in Costa Rica: changes in biomass, nutrient and carbon pools with stand age and plant density. Field Crops Research 74:13-22.

Ares, A., E. Molina, F. Cox, R. Yost, and J. Boniche. 2002b. Fertilizacion fosforada del pejibaye para palmito (*Bactris gasipaes*) en vivero y en plantacion. Agronomia Costarricense 26:63-74.

Ares, A., N. Falcao, K. Yuyama, R. Yost, and C.R. Clement. 2003. Response to fertilization and nutrient deficiency diagnostics in peach palm in Central Amazonia. Nutrient Cycling In Agroecosystems 66:221-232.

Attanandana, T., and R.S. Yost. 2003. A site specific nutrient management approach for Maize: Thailand's experience. Better Crops International 17:3-7.

Attanandana, T., R. Yost, and P. Verapattananirund. 2007. Empowering Farmer Leaders to Acquire and Practice Site-Specific Nutrient Management Technology. Journal of Sustainable Agriculture 30:87-104.

Attanandana, T., P. Verapattananirund, and R. Yost. 2008. Refining and disseminating site-specific nutrient management technology in Thailand. Agronomy for Sustainable Development 28:291-297.

Agus, F., D.K. Cassel, and D. P. Garrity. 1997. Soil-water and soil physical properties under hedgerow systems on sloping Oxisols. Soil Tillage Res. 40:185-199.

Attanandana, T., A. Phonphoen, A. Pinchongskuldit, and R.S. Yost. 2006. SimCorn -- A Software Designed to Support Site-Specific Nutrient Management For Farmers of Small Parcels in The Tropics.

Attanandana, T., C. Suwannarat, T. Vearasilp, S. Kongton, R. Meesawat, P. Boonamphol, K. Soitong, and A. Charoensaksiri. 1999. Nitrogen fertilizer recommendation for grain yield of corn using a modeling approach. Thai J Agric Sci 32:73-83.

Attanandana, T., Russell Yost, and Prateep Verapattananirund. 2004. Adapting site specific nutrient management to small farms of the tropics. International Conference on Precision Agriculture.

Beck, K. 1998. Extreme Programming: A humanistic discipline of software development, p. 1-6, *In* E. Astesiano, (ed.) Fundamental Approaches to Software Engineering. ed. Lecture Notes in Computer Science.

Boonsompopphan, B., T. Vearasilp, R.S. Yost, and T. Attanandana. 2009. Field identification of soil series: Indexing and retrieving soil information while sharing experience and knowledge (vol 173, pg 736, 2008). Soil Science 174:65-65.

Buol, S.W., F.D. Hole, and R.J. McCracken. 1989. Soil Genesis and Classification. 3rd ed. Iowa State University, Ames.

Burges, A.J., and D.P. Lettenmaier. 1975. Probabilistic methods in stream quality management. Water resource bulletin American Water Resource Association 11:115-130.

Chen, G., R.S. Yost, Z.C. Li, X. Wang, and F.R. Cox. 1997. Uncertainty analysis for knowledge-based decision aids: Application to PDSS (phosphorus decision support system). Agricultural Systems 55:461-471.

Colfer, C.J.P. 1991. Toward Sustainable Agriculture in the Humid Tropics: Building on the TropSoils Experience in Indonesia. Bulletin No. 91-02. TropSoils. Raleigh, N.C.

Colfer, C.J.P., R.S. Yost, F. Agus, and S. Evensen. 1989. Expert Systems: A possible link from field work to policy in farming systems. AI Applications in Natural Resource Management 2:31-40.

Colfer, Carol J. Pierce, Dan Gill and Fahuddin Agus. 1988. An Indigenous Agroforestry Model from West Sumatra: A Source of Insight for Scientists. Agricultural Systems 26:191-209.

Deenik, J., A. Ares, and R.S. Yost. 2000. Fertilization response and nutrient diagnosis in peach palm (*Bactris gasipaes*): a review. Nutrient Cycling in Agroecosystems 56:195-207.

Diarra, A.S., R.S. Yost, M.D. Doumbia, T. Attanandana, A. Bagayoko, A.B. Kouyate, R. Kablan, and X.M. Wang. 2004. An algorithm for estimating amounts of rock phosphate needed to meet crop phosphorus requirements in west African soils. Communications in Soil Science and Plant Analysis 35:385-397.

Dierolf, T.S., L.M. Arya, and R.S. Yost. 1997. Water and Cation Movement in an Indonesian Ultisol. Agronomy Journal 89:572.

Dierolf, T.S., L.I. Amien, and R.S. Yost. 1999. Testing and validation of an expert system for making liming decisions. Nutrient Cycling in Agroecosystems 54:9-20.

Evensen, C.L.I. 1989. Alley cropping and green manuring for upland crop production in West Sumatra. Ph.D. Dissertation, University of Hawai`i at Manoa

Fowler, C. 2005. Why is maize a sacred plant? Social history and agrarian change on Sumba. Journal of Ethnobiology 25:39-57.

Fowler, C.T. 1999. The Creolization of natives and exotics: The changing symbolic and functional character of culture and agriculture in Kodi, West Sumba (Indonesia). Ph.D., University of Hawai`i at Manoa, Honolulu.

Fox, R.L., R.S. Yost, N.A. Saidy, and B.T. Kang. 1985. Nutritional Complexities Associated with pH Variables in Humid Tropical Soils. Soil Science Society of America Journal 49:1475-1480.

George, T., R. Magbanua, B. Tubana, J. Quiton, A. Almendras, W. Khatib, F. Cox, and R. Yost. 2000. Estimating buffer coefficients for the phosphorus decision support system: Preliminary comparisons among field and laboratory estimates from three Asian upland soils. Communications in Soil Science and Plant Analysis 31:2101-2110.

Gigou, J., K. Traore, F. Giraudy, H. Coulibaly, B. Sogoba, and M. Doumbia. 2006. Aménagement paysan des terres et réduction du ruissellement dans les savanes africaines. Cahiers Agricultures Vol. 15.

Gill, D.W. 1988. Response of upland crops to potassium at three levels of aluminum saturation in the humid tropics of West Sumatra. Ph. D. Diss. N. Carolina State Univ. Raleigh [Int. B, 1989, 49(7), 2423]

Glaser, R., and M.T.H. Chi. 1988. Overview, *In* M. T. H. Chi, et al., (eds.) The Nature of Expertise. ed. Lawrence Erlbaum Associates, Hilsdale, New Jersey.

Gonzalez-Erico, E., E.J. Kamprath, G.C. Naderman, and W.V. Soares. 1979. Effect of depth of lime incorporation on the growth of corn on an Oxisol of Central Brazil. Soil Sci Soc Am J 43:1155-1158.

Hayes-Roth, F., D.A. Waterman, and D.B. Lenat. 1983. Building Expert Systems. Addison-Wesley Pub. Co., Reading, Mass.

Henschen, D. 2011. Jeopardy Challenge Is More Than A Game. Informationweekcom.26.

Hunington, D. 1985. EXSYS, version 3.2. Expert System Development, Albuquerque, N.M.

Itoga, S.Y., C.L. Chia, R.S. Yost, and R.F.L. Mau. 1990. Propa: A papaya management expert system. Knowledge-Based Systems 3:163-169.

Jones, J.W., G.Y. Tsuji, G. Hoogenboom, L.A. Hunt, P.K. Thornton, P.W. Wilkens, D.T. Imamura, W.T. Bowen, and U. Singh. 1998. Decision Support System for agrotechnology transfer, DSSAT v3, *In* G. Hoogenboom G.Y. Tsuji, P.K. Thornton, (ed.) Understanding Options for Agricultural Production. ed. Kluwer Acadamic Publishers, Dordrecht/Boston/London.

Kablan, R., R.S. Yost, K. Brannan, M.D. Doumbia, K. Traore, A. Yorote, Y. Toloba, S. Sissoko, O. Samake, M. Vaksman, L. Dioni, and M. Sissoko. 2008. "Amenagement en

courbes de niveau," increasing rainfall capture, storage, and drainage in soils of Mali. Arid Land Research and Management 22:62-80.

Kamprath, E.J. 1984. Crop response to lime on soils in the tropics., p. 349-368, *In* F. Adams, (ed.) Soil acidity and liming. ed. American Society of Agronomy, Madison.

Le Istiqlal, A. 1986. Expert Systems. Indonesian Agricultural Research and Development Journal Vol 8 (3&4):72-76.

Li, Hu, Hongbing Kou, Tasnee Attanandana, and Russell Yost. Proc. Computers in Agriculture and Natural Resources, 4th World Congress Conference, Proceedings of the 24-26 July 2006 (Orlando, Florida USA) Publication Date 24 July 2006. 2006.

Li, Z.-C., I.P.G. Widjaja-adhi, T.S. Dierolf, and R.S. Yost. 1995. Liming material selection by computer spreadsheet based on lime and soil characteristics and transportation costs. Journal of Natural Resources and Life Sciences Education 25:26-30.

Michalski, R.S., J.H. Davis, V.S. Bisht, and J.B. Sinclair. 1981. PLANT/DS - An Experimental Computer Consulting System for the Diagnosis of Soybean Diseases. Phytopathology 71:242-242.

Nilawonk, W., T. Attanandana, A. Phonphoem, R. Yost, and X.F. Shuai. 2008. Potassium release in representative maize-producing soils of Thailand. Soil Science Society of America Journal 72:791-797.

Osmond, D.L. 1991. Nitrogen fertilizer recommendations for maize produced in the tropics: Comparison of three computer-based models. Dissertation, Cornell University.

Osmond, D.L., T. Corton, T.J. Smyth, R.S. Yost, and W.S. Reid. 2000. Proceedings of the Workshop on Decision Processes for Determining Diagnostic and Predictive Criteria for Soil Nutrient Management, September 6-10, 1999, PhilRice, . United States Agency for International Development - Soil Management Collaborative Research Support Program (Cornell University, North Carolina State University, Texas A&M University, and University of Hawaii), PhilRice, and the International Rice Research Institute Soil Management Collaborative Research Support Program Technical Bulletin 2000-03. PhilRice, Maligaya, Muñoz, Nueva Ecija, Philippines.

Pearl, J. 1988. Probabilistic Inferences in Intelligent Systems Morgan Kaufmann Publishers, Inc., San Mateo, California.

Rambo, A., Terry, and E. Sajise, Percy. 1984. An Introduction to Human Ecology Research on Agricultural Systems in Southeast Asia East-West Center, Honolulu, Hawaii.

Rich, E. 1983. Artificial Intelligence McGraw-Hill, Inc., New York.

Robotham, M. 1996. Proc. Second Symposium on System Approaches for Agricultural Development, Manila, Philippines. 1996.

Robotham, M. 1998. Staying alive: Sustainability in Philippine upland management systems. Ph.D., University of Hawai`i at Manoa, Honolulu.

Shaner, W.W., P.F. Philipp, and W.R. Schmehl. 1982. Farming Systems Research and Development: Guidelines for developing countries., Boulder, Colorado.

Sidibé, A., R. Yost, M.D. Doumbia, T. Attanandana, A. Bagayoko, A.B. Kouyate, R. Kablan, and X. Wang. 2004. A proposed algorithm for estimating the amounts of rock P needed to meet crop phosphorus requirements in West African Soils. Communications in Soil Science and Plant Analysis 35:385-397.

Sipaseuth, N., T. Attanandana, V. Vichukit, and R. Yost. 2007. Subsoil Nitrate and Maize Root Distribution in Two Important Maize Growing Soils in Thailand. Soil Science 172:861-875.

Stanford, G. 1973. Rationale for optimum nitrogen fertilization of in corn production. J Env Qual 2:159-166.

Stirzaker, R., H. Biggs, D. Roux, and P. Cilliers. 2010. Requisite Simplicities to Help Negotiate Complex Problems. Ambio 39:600-607.

Walker, T., J. Friday, M. Casimero, R. Dollentas, A. Mataia, R. Acda, and R. Yost. 2009. The early economic impact of a nutrient management decision support system (NuMaSS) on small farm households cultivating maize on acidic, upland soils in the Philippines. Agricultural Systems 101:162-172.

Wells. 2001. Extreme programming.

Yampracha, S., T.A. Attanandana, A. Sidibé-Diarra, and R. Yost. 2005. Predicting the Dissolution of Rock Phosphates in Flooded Acid Sulfate Soils. Soil Sci Soc Am J 69:2000-2011.

Yost, R., and T. Attanandana. 2006. Predicting and testing site-specific potassium fertilization of maize in soils of the tropics - An example from Thailand. Soil Science 171:968-980.

Yost, R., A. Ares, J. Bajita, and X. Wang. 1999. Diagnosing nutrient deficiencies, *In* M. Casimero, et al., (eds.) Workshop on diagnostic and predictive criteria in nutrient management. ed. PhilRice, PhilRice, Maligaya, Munoz, Nueva Ecija, The Philippines.

Yost, R., R. Caldwell, M. Constantinides, D. Herbert, and J. Fownes. 1997. The Sustainability of Agriculture and Forestry Production Systems on Soils with Low pH, *In* A. C. Moniz, (ed.) Plant-Soil Interactions at Low pH. ed. Brazilian Soil Science Society.

Yost, R.S., and A. Ares. 2007. Nutrient Management Decision Support Systems for Tree Crops. J of Tropical Forest Science 19:176.

Yost, R.S., A.B. Onken, F.C. S., and Reid. 1992a. The diagnosis of phosphorus deficiency and predicting phosphorus requirement CTAHR, University of Hawaii,, Honolulu, HI 96822.

Yost, R.S., G. Uehara, M. Wade, M. Sudjadi, I.P.G. Widjaja-adhi, and Z.-C. Li. 1986. Expert systems in Agriculture: Determining lime recommendations for soils of the humid tropics. Research Extension Series 089 University of Hawaii, Honolulu, Hawaii.

Yost, R.S., G. Uehara, M. Wade, M. Sudjadi, I.P.G. Widjaja-adhi, and L. Zhi-Cheng. 1988. Expert systems in agriculture: Determining lime recommendations for soils of the humid tropics CTAHR, University of Hawaii,, Honolulu, HI 96822.

Yost, R.S., S. Itoga, Z.C. Li, C. Colfer, L.I. Amien, P. Kilham, and J. Hansen. 1992b. Expert Systems for Soil and Crop Management Information Transfer in Developing Countries, *In* C.Mann, (ed.) Applications of Expert Systems In Developing Countries. ed. Westview Press, Boulder, San Francisco, Oxford.

16

Decision Support Systems in Agriculture: Administration of Meteorological Data, Use of Geographic Information Systems(GIS) and Validation Methods in Crop Protection Warning Service

Racca Paolo, Kleinhenz Benno, Zeuner Thorsten,
Keil Barbara, Tschöpe Beate and Jung Jeanette
ZEPP-Central Institution for Decision Support Systems in Crop Protection
Rüdesheimerstr Kreuznach
Germany

1. Introduction

The mission of the Central Institution for Decision Support Systems (DSS) in Crop Protection (German acronym ZEPP) is to develop, collect and examine existing forecasting and simulation models for important agricultural and horticultural pests and diseases and to adapt these models for practical use. More than 40 weather-based forecasting models for pests and diseases have been successfully developed within the last years. The occurrence of diseases/pests and periods of high-intensity attacks can be calculated with high accuracy. The forecast models are based on different concepts. These range from simple temperature sum models to complex population matrices with integrated rate based algorithms to calculate growth, reproduction and distribution of noxious organisms.
DSS are employed for the

- estimation of disease/pest risk
- estimation of the necessity for pesticide treatments
- forecast of the optimal timing for field assessments
- forecast of the optimal timing for pesticide treatments
- recommendation of appropriate pesticides.

Results of DSS are distributed to the farmers via warning services, using different transmission media (bulletins, letters, faxes and telephone answering machines) and via the internet platform www.isip.de (Information System for Integrated Plant Production) (Röhrig & Sander, 2004). The predictions are suitable for integrated as well as organic farming.
In the following chapter the three basic parts that lead to the creation of a comprehensive and modern DSS for forecasting and warning in integrated crop protection will be analysed. Meteorological data are needed as well as assessed field data as input for decision support systems. With these input data the decision support systems calculate an

output result, e.g. the date of the first appearance of a pest. In the first part of this chapter a software is presented which was developed to administer the data of weather stations and to make it available for prognosis models. In the second part the results of a study how to increase the accuracy of simulation models by using Geographic Information Systems (GIS) are presented. The influence of elevation and geographical location on temperature and relative humidity are interpolated using GIS methods, whereas precipitation data was obtained from radar measurements. These meteorological data were then used as input for the simulation models. The output of the models is presented as spatial risk maps in which areas of maximum risk of a disease are displayed. It is expected that by using GIS methods the acceptance of model outputs will be increased by the farmers. Finally model validation is one of the essential requirements of the model development process to guarantee that models are accepted and used to support decision making. Validation ensures that the model meets its intended requirements in terms of the employed methods and the obtained results. The ultimate aim of model validation is to make the model useful ensuring that the model addresses the right problem, provides accurate information about the system which is being modelled and makes it actually be used. Methods for the validation and evaluation of forecasting models are described in the third part of this chapter.

2. Management of meteorological data

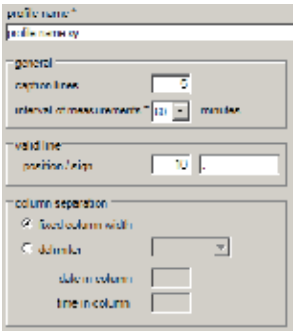
The meteorological data in Germany are provided on the one hand by the German meteorological service, on the other hand some federal states in Germany built up their own meteorological networks. At the moment data of 148 stations of the German meteorological service and 417 stations owned by the federal states are available. In sum these are data of 565 stations which can be used to run decision support systems. The federal states use the locally installed program AgmedaWin (Agrometeorological database for Windows) (Keil & Kleinhenz, 2007) to administer the data of their own weather stations. With AgmedaWin they can import, check, evaluate and store the meteorological data and export them to the internet system www.isip.de in which the data are imported and forecast models are run. AgmedaWin is used so far in 8 states in Germany since 2005. Also the data of the German meteorological service are transferred to www.isip.de.

2.1 Import of meteorological data

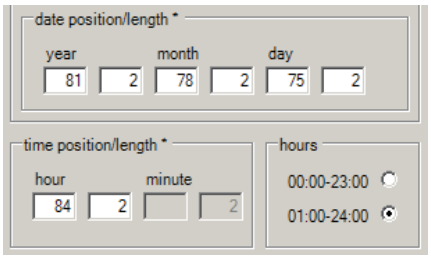
Meteorological data can be imported in AgmedaWin as long as the file containing the meteorological data is an ASCII file. The format of the ASCII files of different stations can vary, depending on the type and the manufacturer and depending on the sensor equipment of the weather station.

With AgmedaWin it is possible to import data having different formats. For this reason an "import wizard" was developed. With the import wizard it is possible to describe exactly the format of almost every ASCII import file by defining "import profiles". The advantage of this solution is that no changes in the program are necessary when a station with a new data format is added. In this case only a new import profile is defined and assigned to the weather station.

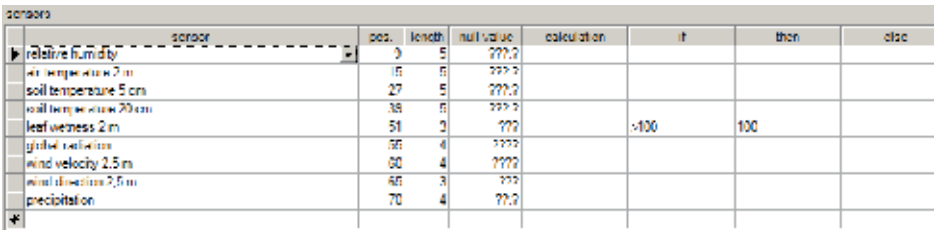
How the format of an ASCII file can be described with an import profile is shown in fig. 1 (a-c).



a. general configurations



b. configuration of date/time format



c. configuration of the sensors format

Fig. 1. AgmedaWin: Definition of an import profile

When defining a new import profile it first has to get a profile name. Then some general configurations can be made by for example entering the number of caption lines and the interval of measurements used in the import file. In AgmedaWin only hourly values are stored. So if the interval of measurements is smaller than 60 minutes the values are automatically aggregated to hourly values before they are imported. Next it is possible to define a valid line. In the example in fig. 1 only values of lines which have a ";" at position 10 will be imported. Also it has to be defined if the sensor values in the import file are arranged in fixed columns or if they are separated by a delimiter (fig. 1.a).

The positions and the lengths of the date and time items as year, month, day, hour and minute have to be entered. It also has to be selected whether in the import file the hours are marked from 00:00-23:00 or from 01:00-24:00 (fig. 1.b).

Also the sensors stored in the import file and the positions and lengths of the sensor values must be entered. In the example in fig. 1.c the value of relative humidity would be expected in column 9 with a length of 5. The format of null values can be entered. Also simple calculations can be done with the values and conditions can be checked before the values are imported (fig. 1.c). After an import profile is defined it is saved and then assigned to a weather station. Only if an import profile is assigned to a weather station data for this station can be imported.

2.2 Gap filling

Decision support systems need complete meteorological data as input. Therefore it is very important to make sure that gaps in the data are filled. In AgmedaWin this can be done in different ways.

Gaps can be filled manually by the weather administrator by simply entering or changing the values.

A gap can also be filled by copying the data of a selected station into the gap.
Small gaps which are not bigger than three hours can be filled automatically by executing a linear interpolation.
Big gaps which are bigger than three hours and not bigger than 15 days can be filled with the data of the best adjacent station. The best adjacent station is found by calculating the correlation coefficient. The data of the station with the best correlation coefficient are then merged into the gap by vertical moving and horizontal rotating so that the tangential points at the beginning and the end fit exactly. This is done by graphical support.

2.3 Plausibility checking

Plausibility checking is a very important feature in AgmedaWin because if the meteorological data provided for the prognosis models are wrong the models will give wrong results.
In AgmedaWin two kinds of plausibility checking are possible.
The first kind is the internal plausibility checking. With this method data of a single station can be checked. Several checking algorithms can be defined and adjusted in AgmedaWin.
Examples for checking algorithms:

- checking of lower and upper limits
- several algorithms for checking the dynamic of the data (e.g.: 12 equal values of air temperature in series would be marked as implausible)
- comparing a value with its previous and following value

When values are found to be implausible by the checking routines they are marked with the plausibility sign "*". The weather administrator has to decide what to do with the marked values. Either he would check if the sensor still works correctly or he could define that a value is plausible although the plausibility checking had marked it as implausible by changing the plausibility sign manually.
The second kind of plausibility checking is the external plausibility checking. With this method the data of adjacent stations can be compared. At first groups of stations in a subregion with similar climatical conditions are defined. Then the deviations of the daily mean values are calculated and shown in a cross table. Deviations which exceed a defined limit are marked. The external plausibility checking can be a help to detect defect sensors.

2.4 Representation of meteorological data

In AgmedaWin meteorological data can be represented in different ways.
The data of all stations can be shown on a map after selecting the sensor, the aggregation (hourly, daily or monthly values) and the date and time.
Data of all sensors of one station can be represented in table form.
Also the values of selected stations and sensors can be represented in a diagram. The user has several possibilities to adjust and configure the representation by selecting or unselecting sensors, by highlighting single sensors or changing the colours. Also it is possible to zoom and scroll in the diagram and to print and save it (fig. 2).

2.5 Evaluation of meteorological data

The following evaluations are available in AgmedaWin:
Sum analysis: Output of the sum of all values of a chosen sensor and station depending on a specified lower and upper limit.

Limit analysis: Showing all hours fulfilling a defined condition.
Evaluations of wind speed and wind direction: Showing a percentaged frequency distribution of wind speed and wind direction values and representing a wind rose of all wind direction values of a specified period.
Long-time mean values: Calculation and output of long-time mean values per month for specified stations and sensors.
Climatic water balance: The climatic water balance is calculated from the evapo - transpiration (Penman, 1948) and the rain fall. The result is represented in a graph or can be stored as a file (fig. 3).

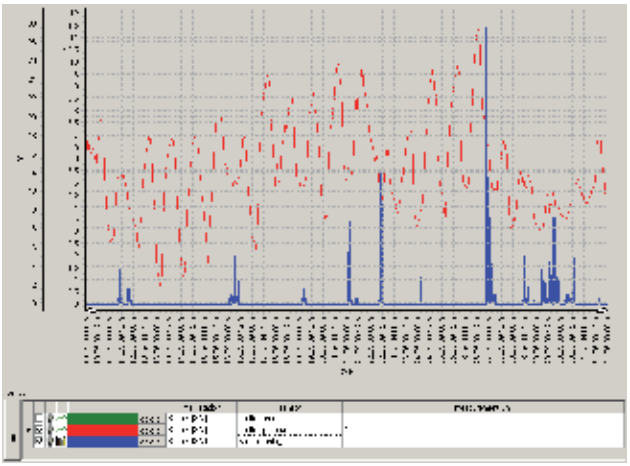


Fig. 2. AgmedaWin: Graphical representation of meteorological data

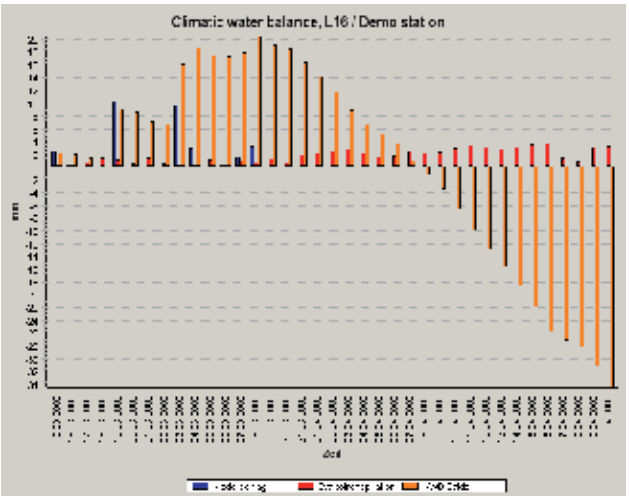


Fig. 3. AgmedaWin: Output of climatic water balance

2.6 Export interfaces

In AgmedaWin an ASCII / Excel export interface is implemented. The user has several possibilities to configure the format of the generated output files. He has to enter the period,

the export format (ASCII or Excel) and the aggregation interval and he has to select the weather stations and sensors. He can also choose whether he wanted the plausibility signs and the minimum and maximum values of the selected aggregation interval to be exported.
The other very important export interface is the export to www.isip.de which consists of two steps. In the first step a XML file is generated containing the information about the data to be exported. This file is zipped and in the second step transferred to www.isip.de via FTP (File Transfer Protocol).
The ISIP export can be done either manually or automatically. If it is done automatically only data which have been changed since the last transfer or which are new are exported to www.isip.de. With this method it is guaranteed that in www.isip.de always the same meteorological data are available as in the local AgmedaWin databases.

3. Use of geographic information systems in crop protection warning service

In the previous chapter the management of meteorological data originating from individual weather stations and their importance for simulation models was highlighted. However, in some agricultural areas, the distance between weather stations (MS) exceeds 60 km. Forecast models did not give satisfactory results for fields separated by such large distances to MSs (Zeuner, 2007).

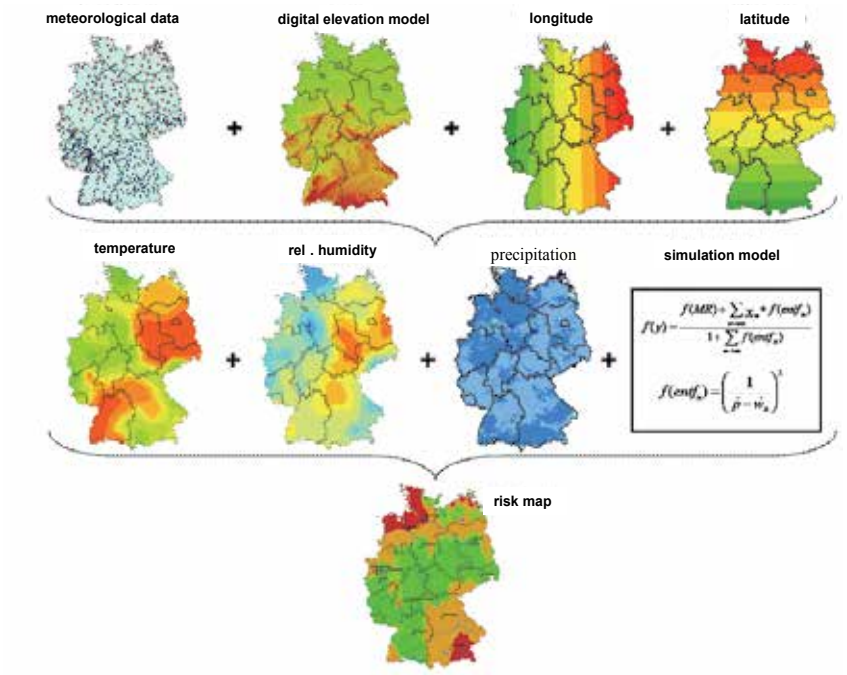


Fig. 4. Process to calculate risk maps using GIS

With the help of Geographic Information Systems (GIS) a plot-specific classification of temperature and relative humidity has been developed using complex statistical interpolation methods described by Zeuner (2007). The method, however, cannot be applied to the parameter precipitation. Especially in the case of frequent spatially and temporally

limited rainfall (so-called convective rainfall event), the interpolation for precipitation does not give plausible results (Zeuner & Kleinhenz, 2007, Zeuner & Kleinhenz, 2008, Zeuner & Kleinhenz, 2009). Precipitation data with a high spatial resolution may be obtained from radar measurements.

Using these spatial input parameters for the currently available disease forecast models should lead to accurate forecasting for areas in-between two or more distant MSs. With the use of GIS, daily spatial risk maps for diseases and pests can be created in which the spatial and the temporal process of first appearance and regional development are documented (Fig.4). These risk maps may lead to improved control and a reduction in fungicide use.

3.1 Storage

In order to store the results of interpolation, a grid was laid out over Germany. At present, the Governmental Crop Protection Services (GCPS) use about 570 MSs to represent an agricultural area of aprox. 200.000 km^2, or an average of one MS per 350 km^2. With the new GIS method, grid cells have a size of 1 km^2 and, after interpolation, are represented by virtual weather stations (Liebig & Mummenthey, 2002)

3.2 Spatial data of temperature and relative humidity

For the interpolation of temperature and relative humidity the multiple regression method was chosen because it gave the best results by the shortest calculation time of all tested interpolation methods. The first calculations with the four interpolation methods (Inverse Distance Weighted, Spline, Kriging and Multiple Regression) showed that deterministic interpolation methods were not suitable. The general purpose of multiple regressions (the term was first used by Pearson, 1908) is to learn more about the relationship between several independant or predictor variables and a dependant or criterion variable. MR is an interpolation method that allows simultaneous testing and modelling of multiple independant variables (Cohen, *et al.*, 2003). Parameters that have an influence on temperature and relative humidity, e.g. elevation, slope, aspect, can therefore be tested simultaneously. MR uses matrix multiplication and only variables with a defined minimum influence that will be included into the model. The result of MR is a formula (x = const + A1*const1 + A2*const2+ A3*const3+…+ Ax*const) which allows a calculation of a parameter set for each grid cell from which independent variables are known (Zeuner, 2007).

year	temperature [°C]				relative humidity [%]			
	2003	2004	2005	2006	2003	2004	2005	2006
CoD	96%	96%	99%	98%	94%	96%	95%	92%
mean dev.	0.0	0.0	0.0	0.1	0.3	0.1	0.1	-0.6
maximum	4.4	4.1	4.3	4.7	19.6	32.6	21.6	21.2
minimum	-3.8	-4.5	-4.5	-4.1	-18.9	-21.9	-22.8	-22.8
t-test	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Table 1. Validation of data on temperature and relative humidity; deviation between calculated values and measured data with MR (n = 92160 hours, n.s. = not significant)

To validate the results of the interpolation, 13 MSs were ignored in the interpolation process. After interpolation, the deviation between calculated values and measured data of these stations was compared. The study was conducted from January to August in the years 2003

to 2006. For all stations, MR gave results with highest accuracy (Tab. 1). In all cases, the coefficient of determination (CoD) ranged between 96 and 99% for temperature and 92 and 96% for relative humidity, respectively. For the 13 MSs, the mean deviation for temperature was less than 0.1°C and for relative humidity less than 0.6% as calculated with MR. The absolute maximum and minimum for temperature was less than 4.7°C and for relative humidity less than 32.6%. The data also were tested for significance between calculated and measured data using a t-test. The test indicated that for all stations the differences between the calculated and measured values were random. The MR method gave plausible results, so it was chosen to interpolate the meteorological data to be used as input for the forecasting models.

3.3 Spatial precipitation data

16 radar stations are run by the German meteorological service to record precipitation all over Germany. These stations do not measure the amount of precipitation at ground level but the signal reflected from the rain drops in the atmosphere. These measurements at first only allowed calculation of an unspecific 'precipitation intensity', a shortcoming. With the system RADOLAN intensity is now calibrated online with data from a comprehensive network of ombrometers, using complex mathematic algorithms. As a result the amount of precipitation can be provided in a spatial resolution of 1 km² (Bartels, 2006). These calibrated amounts of precipitation based on radar measured rainfall intensities are referred to as "radar data" in the following. The validation of precipitation data took place in intensely used agricultural areas, joining the radar grid with stations of the meteorological network. In this way, it was possible to relate each station to a grid cell.

The radar derived precipitation at the station's grid cell and the actually measured data formed the basis for the statistical verification. Since rain events differ throughout the year, two representative months (May and August 2007) were selected to analyse uniform rainfalls in spring as well as convective rainfall events in summer. This resulted in a validation dataset of 1488 hours for each MS. Depending on the region, the number of MSs ranged from 9 to 29. In addition, the influence of the distance between radar station and MSs was analysed.

Furthermore, a leaf wetness simulation model used by ZEPP (Racca, 2001, unpublished) was run on data from both methods of precipitation measurement and the results were compared.

The parameters for the amount of precipitation, number of hours with precipitation and calculated leaf wetness showed high correlations between radar values and measured data. The maximum of the hourly deviation of the amount of precipitation was 0.06 mm. In hours with rainfall the deviation was slightly higher (0.36 mm). No correlation could be detected for the distance between radar stations and MSs. For hourly rainfall pattern, a correlation of 91.4% between stations and validation areas was measured. The best correlations were obtained for the leaf wetness model for which values > 99.9% were achieved.

The results clearly show that the use of radar data as an input parameter in disease forecast models is valid. By adding data of temperature and relative humidity with high spatial resolution, an optimal basis for plot-specific forecasts has been established. Moreover, this system allows the exact detection of local convective rainfall events, which at the moment often remain undetected using individual weather stations. Significant improvements of the spatial forecasting by plant disease simulation models can be expected from the use of radar data.

3.4 Introducing spatial risk maps into practice (www.isip.de)

ISIP, the Information System for Integrated Plant production www.isip.de, is a Germany-wide online decision support system. It has been initiated in 2001 by the German Crop Protection Services as a common portal, thus achieving synergies by pooling existing information. Target groups are farmers as well as advisors.

Since information transfer is the primary task of extension services, the system is intended to make this work more efficient by using modern information technology. Therefore a bi-directional data flow between the services and the farmers was developed. By combining general with specific data, recommendations can be refined from regional to individual. The information is primarily distributed via HTML pages, thus a browser is necessary to use the system (Röhrig & Sander, 2004).

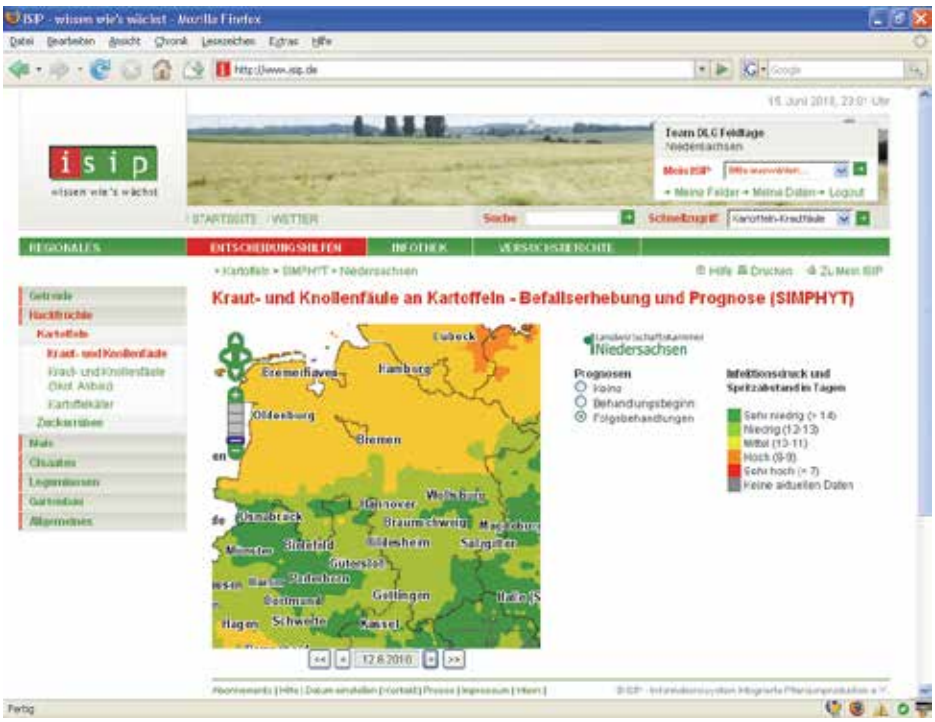


Fig. 5. Risk map of the German Federal State of Lower Saxony for Potato Late blight in mid-June 2010. Shown are the infection pressures in five classes (very low [green] to very high [red]) and the respective spraying intervals in days

In 2010 a new way of presenting results of prognosis models for plant pests and diseases has been implemented. Using interpolated meteorological data in a high spatial resolution as input paramters, so-called 'risk maps' are drawn (fig. 5). These maps have several advantages compared to results representing point information based on single weather stations:

- A risk maps more suitable to identify hot spots and eases the interpretation of the model's results.
- The user does not have to choose a specific weather station, which might even not be valid for his plant production site.

- The maps are produced conform to the OGC standards, thus can be used in other systems.

In addition to the GIS functionalities of zooming and panning, it is possible to scroll through the maps of the last ten days. This gives an excellent overview of the temporal development of the pest or disease risk.
The system is supplemented by a spatial three-day weather forecast 3 day offered by the German Meteorological Service. It is expected that this further supports the decision and management processes of the farmer.

4. Model validation and evaluation in crop protection warning service

Whereas in the previous two chapters the management, validation and preparation of meteorological data which are essential for the forecasting models were discussed the next chapter will address the complex issue concerning the validity of the simulation models used in pest integrated control.
Considering the general mathematical simulation of real events or specific models for plant disease epidemiology, there are several definitions of model validation (Kranz & Royle, 1978, Racca, *et al.*, 2010b, Reynolds, *et al.*, 1981, Sargent, 1998, Schlesinger, 1979, Teng, 1985, Welch, *et al.*, 1981). The validation process can be summarised as the comparison between the virtual (simulated) and the real (actual) system. According to (Balci & Sargent, 1984) the model validation of a generic model will be separated in two methods: the subjective validation techniques and the statistical techniques. In other words, the model validation strictly depends on the modelled system, the model output and the availability of observed field data for the validation.
According to the classification due to ZEPP (Racca, *et al.*, 2010b) the simulation models can be summarised in:
Pests and disease models:

- Type 1: Models predicting first appearance of a disease or a pest.
- Type 2: Complex simulation models predicting the epidemiological development or population dynamics like the disease severity, the disease incidence or pest abundance.
- Type 3: Models predicting a target event like the overriding of an action threshold or periods with high risk for an epidemiological development of the disease or of pest population dynamic.

Ontogenesis models:

- Ontogenesis models are able to simulate the crop-growing. These models can be used in combination with the diseases or corresponding pest models.

For each kind of these models a subjective and/or a statistical validation is possible.
The evaluation of the model performance depends on the aim and the type of model and is strictly correlated with the validation results. Generally, based on the ZEPP experience, a model has a good performance when during the validation phase a quote of correct prognoses up to 75% is overridden and/or the underestimations are less than 15%.

4.1 Validation of type 1 models

Type 1 models forecast the first appearance of a disease or a pest at the beginning of the growing season. They are mainly used by advisors to determine the beginning of the regional monitoring activities and by farmers to do the initial checks on their fields. The results are generally given on a regional level considering the region as the area

surrounding a weather station. Examples of these models currently used by the GCPS are CERCBET1, ERYBET1, UROBET1 and RAMUBET1 used to predict the appearance of Cercospora leaf spot (*Cercospora beticola*) (Roßberg, *et al.*, 2000), powdery mildew (*Erysiphe betae*), rust (*Uromyces betae*) and ramularia (*Ramularia beticola*) (Racca, *et al.*, 2010a) on sugar beet, SIMCOL1 for the lupin anthracnose (*Colletotrichum lupini*) (Racca & Tschöpe, 2010), SIMPHYT1 and SIMBLIGHT1 for the potato leaf blight (*Phytopthora infestans*) (Kleinhenz, *et al.*, 2007), SIMPEROTA1 for blue mold disease of tobacco (*Peronospora tabacina*) (Racca, *et al.*, 2007). An example for pests is the SIMLEP1-Start model to forecast the appearance of the hibernating adults of Colorado potato beetle (*Leptinotarsa decemlineata*) (Jörg, *et al.*, 2007).
All type 1 models can be validated by using statistical and/or subjective validation methods. Some of these methods will be described below with a few examples.
CERCBET1 is a model which is able to forecast the appearance of Cercospora leaf spot on sugar beet fields (Roßberg, *et al.*, 2000, Rossi & Battilani, 1986, Rossi & Battilani, 1991).
The subjective validation of CERCBET1 took place retrospectively with monitoring - data of the years 1995 to 2008 in all German sugar beet growing areas.
The validation was made by comparing

- the date of the first appearance forecasted by the model and observed in the field;
- the forecasted and observed date when 50% of the fields in one region are infected.. This date represents the distribution of the disease in several fields in one region, the 50th percentiles. At this point the probability to detect a Cercospora infection in a field is very high and the disease had been established in this region. For the validation the available data are grouped in "regions" near a representative weather station. To detect the distribution of the infected field, only regions with surveys in more than four sugar beet fields were taken into consideration.

In any case the forecasting was considered:

- correct - when the difference between the forecasted and the observed date was in a range of one week (± 7 days);
- early/late - when the difference between the forecasted and the observed date was bigger than one week (> ± 7 days).

The subject of the validation method in this case is to consider a period of ± 7 days correct for these kind of model results. We consider that the data for the validation derive from regional monitoring arranged by the GCPS. This monitoring is done weekly. So, one week of delay or one week of earlier forecast is acceptable for this model.
The result of the validation is summarised in Table 2.

	first appearance	50 % infected fields
too early	31.80%	20.14%
correct	64.66%	72.08%
too late	3.53%	7.77%

Table 2. CERCBET1: Subjective validation in Germany. Data from 1995 to 2008 (n=283)

Concerning the first appearance, in about 64% of the forecasts the model was able to predict the disease appearance correct, in 32% the date was too early and in 4% of the cases the forecasted date of disease appearance was too late. The model shows more accuracy in the prediction of the 50% infected field date. Analyzing the results of the validation, the trend of the model to anticipate the occurrence of disease can be identified (Tab.2). This trend can

also be explained by analyzing the same data which was used for the validation. Sometimes sample size t used in the surveys is not appropriate for detecting a rare event like the appearance of first necrotic spots (Roßberg, *et al.*, 2000). It can also be difficult to recognize the first symptoms at the leaves (the first spots of Cercospora could be taken for *Alternaria sp.*, *Phoma sp.* or bacterial spots).

For an appropriate statistical validation (Racca, *et al.*, 2010b, Rossi, *et al.*, 1997a, Teng, 1981) simulation and field data were regarded as two independant random samples to compare the distribution. It is possible to apply some parametric tests like the t-test (comparison of mean values) and the F-test (comparison of standard deviation), but also a non-parametric method like the Kolmogorov-Smirnov test (computing the maximum distance between the cumulative distributions of two samples) (Tab.3)

year	n	first appearance			50% infected fields		
		t-test	F-test	Kol.Smirn. test	t-test	F-test	Kol.Smirn. test
1999	25	n.s.	*	*	n.s.	*	*
2000	16	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2001	16	n.s.	n.s.	*	n.s.	n.s.	n.s.
2002	27	n.s.	*	*	n.s.	n.s.	*
2003	30	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2004	22	n.s.	n.s.	*	n.s.	n.s.	*
2005	35	*	*	*	*	*	*
2006	36	*	n.s.	*	*	n.s.	*
2007	28	*	*	*	n.s.	*	*
2008	28	n.s.	n.s.	*	n.s.	n.s.	n.s.

Table 3. CERCBET1: Statistical tests on the results for the simulation years 1999-2008 (Kol.Smirn.: Kolmogorov-Smirnov test, n.s. not significant, * = significant with $p<0,05$)

Considering the statistical analysis of model results, it must be concluded that the presence of significant differences between the distributions does not show a good correlation according to the data. However, analyzing the subjective validation method we conclude that the model is satisfactory if we consider the principal aim of the model, determine the date of beginning of the monitoring system. An early forecast can also be accepted (in only 12% of the total cases the forecast was more than 3 weeks before the observed first appearance).

A similar validation was done for the SIMBLIGHT1 model. The model predicts the risk of a potato late blight outbreak and recommends the date for the first treatment. A subjective validation was done and the model results were considered valid when the predicted date of the first treatment was earlier than the date of the field observation. More than 700 observations during the period 1994-2005 were recorded for the validation. For most years the proportion of correct forecasts reached more than 90% and a statistical validation was not done (Kleinhenz, *et al.*, 2007).

For models that provide results in a binary response it is possible to apply a different statistical method of evaluation and validation. As one example in the model ERYBET1 the dates of onset of the disease (powdery mildew on sugar beet) were classified into two groups: early onset (before July 31) and late onset (after July 31). The two groups are discriminated by a binary logistic regression model using the winter weather as input parameters (Racca, *et al.*, 2010a).

Binary logistic regressions results in a value between 0 and 1. Discrimination of individual cases in two classes is done by definition of a "cut-off" threshold. Depending on the "cut-off" the following classification of the cases is defined:

- true negative: cases correctly classified in early onset (model specificity)
- false negatives: the values of early onset were classified in late onset (underestimation)
- true positive: cases correctly classified in late onset (model sensitivity)
- false positives: the values of late onset were classified in early onset (overestimation)

The cut-off value is simply chosen mathematically or graphically so that the rate of false positives and false negatives is minimized (Hadjicostas, 2006). The model validation is done using a ROC (Receiver Operating Characteristic) curve (Madden, *et al.*, 2008).

In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (1-Specificity) for different cut-off points of a parameter. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve is a measure of how well a parameter can distinguish between two diagnostic groups (Fig. 6).

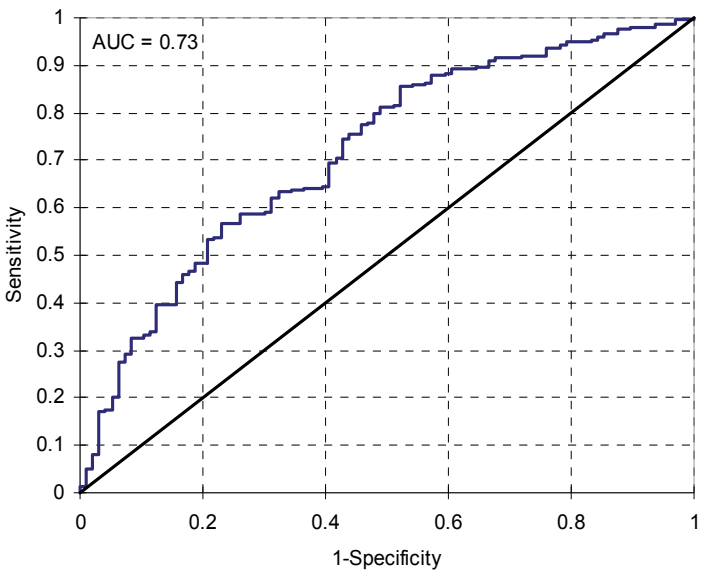


Fig. 6. ROC Curve for the Model ERYBET1 discriminates an early or late powdery mildew onset in sugar beet

Accuracy is measured by the area under the ROC curve (AUC). An area of 1 represents a perfect test, an area of 0.5 represents a worthless test. A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system: 0.9-1 = excellent; 0.8-0.9 = good, 0.7-0.8 = fair, 0.6-0.7 = poor, 0.5-0.6 = fail.

Since the AUC of the ERYBET1 model is equal to 0.73 it is statistically acceptable. Furthermore, with a cut-off equal to 0.74 (estimated to obtain for the model a weighted combination of specificity and sensitivity) the values shown in Table 5 are obtained for the classification of the onsets. The percentage of correct classification is about 65% of all cases. The underestimations were less than 35% and so the model is considered a good estimator of the early or late onset of disease dependant only on winter weather conditions.

disease onset classification	early	late	sum	correct (%)
early	54	39	93	58.06%
late	80	164	244	67.21%
sum	134	203	337	64.69%

Table 4. Results of the classification of the disease onset (powdery mildew on sugar beet) forecasted by the model ERYBET1 using binary logistic regression with "cut off" value = 0.74

4.2 Validation of type 2 models

Type 2 models are examples for classical simulation models. Generally, they are very complex. The aim is to predict the epidemic development (expressed as disease severity and/or disease incidence) for the diseases or to predict the various development stages of insects. These models are used, in the examination phase, as a basis for identifying the parameters and variables for the construction of type 1 and 3 models. Examples of these models are CERCODEP which is able to simulate the epidemics of Cercospora leaf spot on sugar beet (Rossi, *et al.*, 1994), RUSTDEP for leaf rust (*Puccinia recondita*) on winter wheat (Rossi, *et al.*, 1997b), SEPTRI2 for leaf blotch (*Septoria triciti*) on winter wheat (Erven, *et al.*, 2008, Erven, *et al.*, 2010) , SIMPHYT2 for the potato leaf blight (Roßberg, *et al.*, 2001) and SIMLEP2 to forecast the phenological development of the Colorado potato beetle (Roßberg, *et al.*, 1999).

The following three models are used as examples for validation methods: PUCREC for leaf rust in winter rye (*Puccinia recondita*), PUCTRI for leaf rust in winter wheat (*Puccinia triticina*) and SEPTRI for leaf blotch (*Septoria triciti*) in winter wheat.

The first two models simulate the development of the epidemics of rust on different leaf layers expressed as disease incidence (Racca, *et al.*, 2008, Räder, *et al.*, 2007). Models were validated with both subjective and statistical methods of field data, collected from 2002 to 2005. In total, 51 data sets for PUCREC and 37 for PUCTRI were available to investigate the predictive ability of the models.

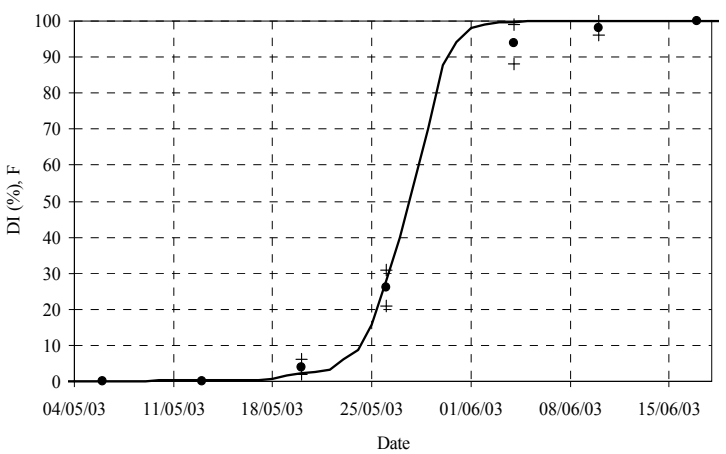


Fig. 7. PUCREC – Simulated disease incidence (DI) on the flag leaf (F). weather station Herxheimweyher (Rhineland-Palatinate – Germany) in 2003. (▬ simulation, • field data, + confidence interval of the field data)

The subjective validation simply consists of the comparison of the simulated disease incidence with data, recorded in the field (fig. 7). The simulation is decided to be correct when the simulated disease incidence ranges between the confidence interval of the recorded disease incidence. Overestimation is given when the simulated value overrides the highest level of the confidence interval in opposition to underestimation when the simulated value is under the lowest level of the confidence interval. Validation was done for both models and for each leaf layer F (flag leaf) to F-3 (Tab.5).

leaf layer	PUCREC -winter rye			PUCTRI -winter wheat		
	under.	corr.	over.	under.	corr.	over.
F	8	74	18	0	82	18
F	8	74	18	0	82	18
F-1	2	86	12	0	76	24
F-2	6	84	10	0	71	29
F-3	0	80	20	0	76	24

Table 5. Validation of PUCREC (n=51) and PUCTRI (n=37) – Share (%) of underestimated, correct and overestimated leaf rust epidemics on different leaf layers (2001 – 2005) (under.= Underestimation, corr. = correct, over.=overestimation)

According to the subjective validation the field data and the model results coincided well. In most of the cases (from 71 to 86%), the disease incidence progress was correctly simulated (Tab. 5). In a few cases PUCREC underestimated (2 - 8%) or overestimated (10 – 20%) the epidemic progress of *P. recondita*. For winter wheat a considerable share of overestimations occurred (18 – 29%). This means that epidemics simulated by PUCTRI started earlier and progressed faster than observed in the field. For winter wheat no underestimations could be observed.

The statistical validation was done with two parametric (regression analysis, hypothesis test) and one non-parametric test (Kolmogorov-Smirnov).

The simulated disease incidence (dependant variable) is simply linear correlated with the recorded data (independant variable). The "null hypothesis" demonstrates that "a" (intercept of the regression line) is equal to 0 and "b" (slope of regression) is equal to 1 (tested using the Student t-test) (Tab.6).

Leaf layer	PUCREC -winter rye						PUCTRI -winter wheat					
	Regression parameters				Kolm.-Smirn.		Regression parameters				Kolm.-Smirn.	
	t-a		t-b				t-a		t-b			
	ns	*	ns	*	ns	*	ns	*	ns	*	ns	*
F	93	7	59	41	96	4	95	5	90	10	95	5
F-1	91	9	77	23	98	2	94	6	87	13	94	6
F-2	91	9	68	32	96	4	100	-	100	-	92	8
F-3	88	12	79	21	91	9	100	-	67	33	100	-

Table 6. Validation of PUCREC (n=51) and PUCTRI (n=37) – regression analysis and Kolmogorov-Smirnov test (2001 – 2005), share (%) of the significance. (t-a: hypothesis t-test for regression intercept, t-b: hypothesis t-test for regression slope, Kol.Smirn.= Kolmogorov-Smirnov test, n.s. = not significant, * = significant with $p<0.05$)

The statistical validation gave very satisfactory results with both (parametric and non-parametric) methods. The high number of non-significant cases of the regression parameters

and the Kolmogorov-Smirnov test mean that the model is considered a statistically accurate simulator of the field data (Teng, 1981).
SEPTRI is a model able to simulate the disease epidemics of *Septoria tritici* on winter wheat. In particular the model calculates the infection probability, the infection rate, the lesion growth and the sporulation for wheat cultivars for three susceptibility levels: high susceptible, mean susceptible and low susceptible. The simulation results are the values of disease severity per leaf layer.
Also this model was validated using both subjective visual method (fig. 8, tab. 7) and statistical methods (tab. 8).

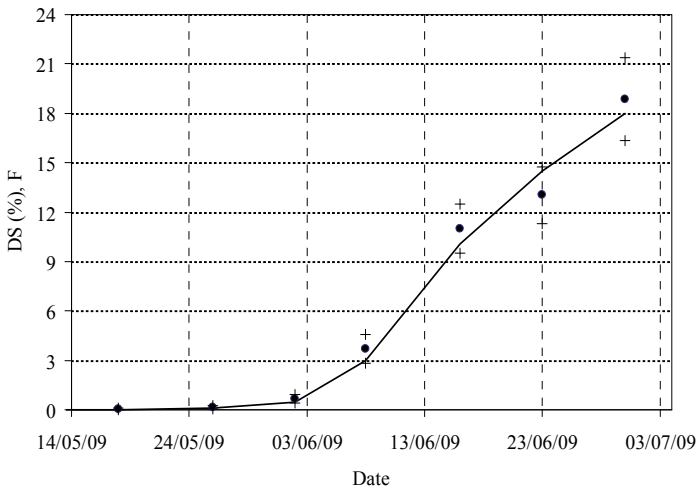


Fig. 8. SEPTRI – Simulated disease severity (DS) on the flag leaf (F). weather station Rommersheim (Rhineland-Palatinate – Germany) in 2009. Wheat cultivar with high Septoria susceptibility (▬ simulation, • field data, + confidence interval of the field data)

leaf layer	cultivar susceptibility								
	high			mean			low		
	under.	corr.	over.	under.	corr.	over.	under.	corr.	over.
F	6.3	81.3	12.5	6.3	75.0	18.8	6.3	62.5	31.3
F-1	31.3	68.8	0	12.5	81.3	6.3	6.3	68.8	25.0
F-2	50.0	50.0	0	31.3	68.8	0	12.5	68.8	18.8
F-3	75.0	25.0	0	62.5	37.5	0	43.8	50	6.3

Table 7. Validation of SEPTRI (n=60) – share (%) of underestimated, correct and overestimated leaf blotch Epidemics on different leaf layers (2007 – 2009) (under.= underestimation, corr. = correct, over.=overestimation)

According to the subjective validation, the model results are satisfactory. In most of the cases (from 50 to 81.3%), excluding the leaf layer F-3, the disease severity progress was simulated correctly (Tab. 7). For leaf layer F-3 a considerable share of underestimation occurred (from 43.8 to 75%). This means that epidemics on this leaf layer were poorly simulated. Otherwise for use of the model in praxis the leaf layer F-3 has less importance because the action threshold for the treatment scheduling is generally based on leaf layer F-2.

leaf layer	SEPTRI – all cultivar susceptibilities					
	regression parameters				Kolm.-Smirn.	
	t-a		t-b			
	ns	*	ns	*	ns	*
F	100	0	87	13	74	26
F-1	100	0	100	0	95	5
F-2	100	0	90	10	65	35

Table 8. Validation of SEPTRI (n=60) – Regression analysis and Kolmogorov-Smirnov test (2007 – 2009), share (%) of the significance. (t-a: hypothesis t-test for regression intercept, t-b: hypothesis t-test for regression slope, Kol.Smirn.: Kolmogorov-Smirnov test, n.s. not significant, * = significant with $p<0.05$)

The statistical validation of the SEPTRI model gives satisfactory results for the leaf layer F, F-1 and F-2 with both (parametric and non-parametric) methods.

4.3 Validation of type 3 models

Type 3 models are derived from models of type 2. The output is various. The simulation of the development of the disease is often done to forecast the overriding of the action threshold and to give a recommendation for a fungicide or insecticide spraying. Models of type 3 are sometimes combined with routines for the calculation of the effectiveness of pesticides. They may include agricultural parameters like type of rotation, fertilization, irrigation and cultivar susceptibility which interact with the epidemics of the specific disease. The type 3 models can be used on regional and field-specific levels. Some examples of the most successful type 3 models are CERCBET3 (Racca & Jörg, 2007), ERYBET3, UROBET3 and RAMUBET3 (Racca, *et al.*, 2010a) for Cercospora leaf spot, powdery mildew, rust and ramularia leaf spot respectively on sugar beet, SIMPHYT3 for potato leaf blight (Gutsche, 1999), PUCREC3 and PUCTRI3 for cereal leaf rusts (Räder, *et al.*, 2007) and SIMLEP3 for Colorado potato beetle (Jörg, *et al.*, 2007).

Some validation methods for type 3 models can be described using sugar beet leaf disease models as examples. CERCBET3, ERYBET3, UROBET3 and RAMUBET3 simulate the progress of sugar beet leaf diseases, expressed as disease incidence development, and forecast the overriding of an action threshold, suggested for the treatments during the sugar beet growing season (Racca & Jörg, 2007, Racca, *et al.*, 2004). For most sugar beet growing areas in Germany the Regional Plant Protection Offices use an action threshold for the fungicide treatment to control sugar beet leaf diseases based both on time and disease incidence (DI). Particularly the Plant Protection Services suggest the following strategy: action threshold 5% DI until the end of July, 15% DI before the 15th of August and 45% DI later than 15th August (Jörg & Krauthausen, 1996). For a subjective validation the weekly assessment data of the disease incidence were confronted with simulated data. The difference between the assessed and simulated date exceeding the relevant action threshold (5, 15 and 45% of disease incidence) was classified as follows:

- too early: the forecast exceeds the action threshold more than 7 days earlier than the assessed date;
- correct: the forecast differs not more than ± 7 days from the assessed date;
- too late: the forecast exceeds the action threshold more than 7 days later than the assessed date.

The results of this subjective validation is summarised in Tab. 9.

model	n	too early	correct	too late
CERCBET3	71	10.9 %	84.2 %	4.9 %
ERYBET3	555	28.4 %	63.8 %	2.9 %
UROBET3	652	16.5 %	77.2 %	6.3 %
RAMUBET3	241	9.6%	86.0 %	2.5 %

Table 9. Sugar beet type 3 models validation. Comparison of the assessed and the forecasted date for exceeding the action threshold (year 2001-2009)

All sugar beet type 3 models gave very satisfactory results. The action threshold was forecasted correctly in about 84% of the cases for cercospora leaf spot, in about 69% for powdery mildew, in about 77% of the cases for rust and in about 86% of the cases for ramularia leaf spot.
For the CERCBET3 model the same pool of data was useful for an appropriate statistical validation. For all three action thresholds the simulated date of threshold exceeding was linear regressed with the real date (Tab.10).

threshold	a	b	t-a	t-b	r^2	ρc
action threshold 5%	5.43	0.86	*	*	0.72	0.82
action threshold 15%	20.76	0.31	*	*	0.31	0.45
action threshold 45%	32.64	0.24	*	*	0.19	0.36

Table 10. Parameters of the regression analysis between the simulated and the real date of threshold overriding (t-a: hypothesis t-test for regression intercept, t-b: hypothesis t-test for regression slope, n.s. not significant, * = significant with $p<0.05$, ρc: concordance correlation coefficient).

The statistic analysis shows different results compared to the subjective validation. There is only a strong statistical correlation for the action threshold 5% ($r^2 = 0.72$). The other action thresholds are poor correlated ($r^2=0.31$ and 0.19). For all regressions the intercept and the slope are significant at $p<0.05$.
One step further, using the concordance correlation coefficient ρc (Lin, 1989) can be overcome the statistical validation of avoiding problems with misinterpreted results in the regression analysis (failed t-test for a and b). The values of ρc range from 1 perfect agreement to -1, total lack of agreement. Validation of the action threshold 5% shows a high value of ρc. There is a good agreement between the simulated and the real observations.
In practical use of the model, a correct timing of the overriding of the first threshold (action threshold 5%) is very important. In some cases a treatment in this early phase is able to decelerate the epidemic of the disease and a second or third spraying can be avoided. In this case the combination of subjective and statistical validation is very satisfactory. Another example for the validation of a type 3 model can be illustrated with the model SIMLEP3. The model simulates the development of *Leptinotarsa decemlineata* from the beginning of egg laying to the occurrence of the old larvae at a field specific scale (Jörg, *et al.*, 2007). SIMLEP3 was validated in Germany and in several European countries. The method is only

subjective, comparing the forecasting dates of the maximum abundance of egg cluster and young larvae with field observations. The model output was considered correct when the forecast was within an interval of one week compared to the observed date (Tab. 11).

country	maximum abundance of					
	egg clusters			young larvae		
	% correct	% too early/late	n	% correct	% too early/late	n
Germany	91	9	33	87	13	38
Italy	100	0	6	100	0	6
Austria	71	29	7	86	14	7
Poland	100	0	2	100	0	2
Mean	90.5	7.5		93.25	6.75	

Table 11. Results of SIMLEP3 subjective validation in several European countries (1999-2004): share of correct forecasts (%).

In general, SIMLEP3 results were very satisfying. The first occurrence of young larvae in most of the cases was predicted correctly. Nevertheless, differences between forecasting and observed date ranging from 18 days too early up to 10 days too late were registered. Good results were also obtained for the prediction of maximum egg cluster occurrence. Throughout Germany, Poland, Austria and Italy the mean share of correct forecasts given from SIMLEP3 (both egg clusters and young larvae) amounted to about 92%. In Austria, the share of correct predictions was the lowest (approx. 70%) and in Germany the share of correct predictions exceeded 90%. Maximum occurrences of young larvae predictions were correct in about 93% of the cases on the European scale. Again, optimum results were obtained in Italy and Poland. In Austria and Germany, the share of correct forecasts exceeded 85%. Subjective validation efforts showed that SIMLEP3 is able to give correct forecasts for the most important development stages of *L. decemlineata* with respect to control efforts. The validation also demonstrates the possibility to expand the use of the model throughout Europe.

4.4 Validation of ontogenesis models

Ontogenesis models are ontogenetic models which simulate the development of crops expressed as BBCH growth stages (Hack, *et al.*, 1992) over time. SIMONTO-models are based on the modelling approaches of CERES-Wheat (USA) (Gabrielle, *et al.*, 1998) and ONTO-models (Germany) (Wernecke & Claus, 1996). The ontogenetic progress in SIMONTO is reflected by a developmental rate which is a function of temperature and photoperiod. Parameters for the different models (winter oilseed rape and winter cereals) were estimated by employing the Monte-Carlo-method (Falke, *et al.*, 2006, Roßberg, *et al.*, 2005). More than 13800 single observations of BBCH growth stages for winter cereals from 2003 to 2008 were available for the model validation. In the first step of the statistical validation, the observed BBCH growth stages are simply linear regressed with the model. A high r^2 (0.88) suggests a good correlation between the data. Both regression parameters, a and b, are significant but the concordance correlation coefficient of 0.92 demonstrates a good agreement between the data. The model is apparently good to simulate the reality. Unfortunately, the BBCH growth stages are not strictly arithmetical dependant. Some stages could appear very early in the season and stay constant for a long time. The simple arithmetic difference between two BBCH growth stages can be minimal but the difference in

days between the two stages themselves can be very big. For example, the arithmetical difference between BBCH 21 (beginning of the tillering) and 22 (2 tiller detectable on the plants) is only = 1 but sometimes BBCH 21 is recorded on the fields in autumn and simulated by the model in spring. This can mean a difference of up to 5 months. Again the model should also be validated with a subjective method. For SIMONTO a scoring model approach is used (Roßberg, *et al.*, 2005). The difference in days between the simulation and the observation is classified with a subjective flaw dot table:

- model more than 7 days too early or more than 7 days too late: flaw dot 7;
- model too early or too late (4 to 7 days): flaw dot 3;
- model too early or too late (1 to 3 days): flaw dot 1;
- no difference between simulation and observation: flaw dot 0;

The sum of the flaw dots could be classified in a flaw dot coefficient with values varying from 0 (perfect model) to 7 (simulation extremely early or extremely late). Concerning the results in Tab. 12 the flaw dot coefficient is comprised in a range from 1.27 to 3.10. Values are classified when the simulation results arrive with a delay or an advance of up to 3 days. Values above 3 indicate a higher difference of days between simulation and reality.

Since most of the flaw dot coefficients are lower than 3 we can conclude that the model accurately simulates the reality with a gap of acceptable days. In this case, the subjective validation is essential because the statistical validation could lead to misleading results.

BBCH	n	case with				flaw dot	flaw dot coefficient
		flaw dot 7	flaw dot 3	flaw dot 1	flaw dot 0		
BBCH 23	22	4	2	0	18	28	1.27
BBCH 32	1490	252	228	320	636	2930	1.97
BBCH 39	1070	140	203	188	514	1852	1.73
BBCH 61	703	216	212	60	224	2181	3.10
BBCH 65	732	184	0	142	194	2066	2.82

Table 12. SIMONTO: Validation. Error point, sum of the flaw and flaw dot coefficient for some BBCH growth stages in winter cereals (seasons 2003-2008).

Another type 4 model validated in 2010 (Tschöpe & Racca, 2010) was SIMONTO-lupin, a model which simulates the crop stages of the lupin (*Lupinus angustifolius*).

For the subjective validation the simulated progress of ontogenesis was compared visually with the observed BBCH growth stages in the field. The classification of the cases in correct, overestimated and underestimated was the same used for the type 2 model validation. In total 215 data sets were analysed by this validation method. 88.4% of the cases were simulated correctly, 9.8% were underestimated and 1.9% were overestimated.

For the statistical validation the model output (dependant variable) was compared to the field data (independent variable) with the help of a regression analysis. The average of the coefficient of determination was 0.984 and the slope b was in all cases not significant.

BBCH stage	n	too early	correct	too late
61	229	0 %	86.0 %	14.0 %
69		22.7 %	75.5 %	1.7 %

Table 13. Evaluation of the deviation in days (% of the cases with early, correct or late classification) between the date of the BBCH stage observed in field and the simulated for BBCH 61 (begin of flowering) and BBCH 69 (end of flowering) (n=229).

Like the BBCH growth stages of the cereals also the BBCH stages of the lupin are not strictly arithmetically dependant. Also in this case another validation criterion was needed. Again the deviation in days between the ontogenesis in field and the simulated ontogenesis was compared. If the deviation between the both dates was ±7 days the prognosis was rated as "correct". Otherwise the prognosis was too early or too late (Tab. 13).
In total 229 data sets were analysed by this validation method for the BBCH-stages: Start of flowering (BBCH 61) and end of flowering (BBCH 69). BBCH 61 achieved a hit rate of 86.0% correct forecasts, 14.0% of the dates were simulated too late. Concerning the BBCH 69 it was simulated correctly in 75.5% of all cases, 22.7% of the simulations were too early and in 1.7% of the cases the simulated BBCH was too late.

5. Conclusions

Decision support systems in plant protection need plausible and complete meteorological data as main input. Meteorological data on the one hand are provided by the German meteorological service. On the other hand several states in Germany built up their own meteorological networks. These states use the software AgmedaWin for import, management, presentation, evaluation and export of the measured data. Core of the program is a flexible import module which facilitates the import of files with different formats from all types of weather stations by describing the structure of the files with import profiles. Several algorithms are integrated in AgmedaWin to ensure plausibility and completeness of the data. The program also includes a module to compare data of corresponding stations. With an XML-based export interface the data are transferred from AgmedaWin to the internet system www.isip.de where all data are stored and used as input for the decision support systems. Furthermore the unprocessed meteorological data can be evaluated in www.isip.de or downloaded as files in different formats by external users.
The plausibility and completeness of meteorological data as main input for the models is the most important pre-condition to get correct prognosis results. However by using meteorological data of weather stations a good prognosis is only reached in the scope of a weather station. That is the reason why the ZEPP developed a new technology based on Geographic Information Systems (GIS). With the help of GIS it is possible to obtain results with higher accuracy for disease and pest simulation models. The influence of geographical factors on temperature and relative humidity were interpolated with GIS methods getting meteorological data for every km^2 in Germany. The parameter precipitation was taken by radar measured precipitation data and the results of all measured meteorological data were used as input for the simulation models. The output of these models is presented as spatial risk maps in which areas of maximum risk of the disease outbreak, infection pressure or pest appearances are displayed. The modern presentation methods of GIS lead to an easy interpretation and will furthermore promote the use of the system by farmers.
Finally the validation of a simulation model is a critical point in the development of the model itself. Unfortunately, there is no set of specific tests or decision-making algorithms which can determine the best method to validate a model. The subjective methods are certainly more intuitive and provide easy answers with easy interpretations. In this case, the decision for the method depends on the experience of the one who validates the model. It is important to know, for example, what weight has to be indicated to the over- and especially to the underestimation of the results of the model. Careful attention must be payed to the quality of data which is available for the validation. They should certainly be adequate in

number and represent the real agronomical system. Unfortunately there is no "manual for the validation of a model" but a combination of statistical and subjective methods gives good judgement of benefits of a decision support system which helps to plan crop protection activities like field assessments or pesticide use.

6. References

Balci, O., Sargent, R. G. (1984). A bibliography on the credibility assessment and validation of simulation and mathematical models. *Simuletter*, 15, 15-27.

Bartels, H. (2006). Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer).Offenbach.

Cohen, J., Cohen, P., West, S., Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences*, Lawrence Erlbaum, 0805822232.

Erven, T., Jörg, E., Kleinhenz, B., Racca, P., Hau, B. (2008). Untersuchungen zur Epidemiologie von *Septoria tritici* in Abhängigkeit von der Winterweizensorte (Poster). *56. Deutsche Pflanzenschutztagung in Kiel 22.-25. September 2008. Pflanzenproduktion in Wandel - Wandel im Pflanzenschutz?*, J. Kühn-Institut, P.-T. Deutsche, eds., Julius Kühn-Institut, Kiel, 415.

Erven, T., Racca, P., Kleinhenz, B. (2010). *Septoria tritici* an Winterweizen – Das witterungsbasierte sortenabhängige Prognosemodell SEPTTRI1. *57. Deutsche Pflanzenschutztagung. 6. - 9.September 2010 Humboldt-Universität zu Berlin; Gesunde Pflanze, gesunder Mensch*, D. Pflanzenschutztagung, J. Kühn-Institut, eds., Julius Kühn-Inst., Bundesforschungsinst. für Kulturpflanzen, Berlin, 258.

Falke, K., Jörg, E., Roßberg, D. (2006). Ursachen für Abweichungen zwischen bonitierten und simulierten Entwicklungsverläufen von SIMONTO für Winterraps und Wintergetreide *55. Deutsche Pflanzenschutztagung in Göttingen vom 25.-28. September 2006*, BBA, Göttingen, 282.

Gabrielle, B., Denoroy, P., Gosse, G., Justes, E., Andersen, M. N. (1998). Development and evaluation of a CERES-Type model for winter oilseed rape. *Field Crops Res.*, 57, 95-111.

Gutsche, V. (1999). Das Modell SIMPHYT 3 zur Berechnung des witterungsbedingten Epidemiedruckes der Krautfäule der Kartoffel (*Phytophthora infestans* (Mont.) de Bary). *Nachrichtenbl. Deut. Pflanzenschutzd.*, 51(7), 169-175.

Hack, H., Bleiholder, H., Buhr, L., Meier, U., Schnock-Fricke, U., Weber, E., Witzenberger, A. (1992). Einheitliche Codierung der phänologischen Entwicklungsstadien mono- und dikotyler Pflanzen. Erweiterte BBCH-Skala, Allgemein-. *Nachrichtenbl. Deut. Pflanzenschutzd.*, 44(12), 265-270.

Hadjicostas, P. (2006). Maximizing proportions of correct classifications in binary logistic regression. *Journal of Applied Statistics*, 33(6), 629 - 640,0266-4763.

Jörg, E., Krauthausen, H. J. (1996). Bekämpfung von *Cercospora beticola* an Zuckerrüben. *Mitt. Biol Bundesanst. Land - Forstwirtsch*, 321, 185.

Jörg, E., Racca, P., Preiß, U., Butturini, A., Schmiedl, J., Wójtowicz, A. (2007). Control of Colorado Potato Beetle with the SIMLEP Decision Support System. *EPPO Bulletin*(37), 353-358.

Keil, B., Kleinhenz, B. (2007). AgmedaWin – a tool for easy and flexible management of meteorological data. *OEPP/EPPO*(37), 335-338.

Kleinhenz, B., Falke, K., Kakau, J., Roßberg, D. (2007). SIMBLIGHT1 - A new model to predict first occurrence of potato late blight. *Bulletin OEPP/EPPO Bulletin*(37), 1-5.

Kranz, J., Royle, D. J. (1978). *Perspectives in mathematical modelling of plant disease epidemics*, Plant disease epidemiology, Oxford.

Liebig, W., Mummenthey, R. D. (2002). *ArcGIS-ArcView(c)*, Halmstad: Norden.

Lin, L. L. K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45, 225-268.

Madden, L., Hughes, G., Van Den Bosch, F. (2008). *The Study of Plant Disease Epidemics*, The American Phytopathological Society, 978-0-89054-354-2, St. Paul, Minnesota.

Penman, H. L. (1948). Natural Evaporation from Open Water, Bare Soil and Grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 193(1032), 120-145.

Racca, P., Jörg, E. (2007). CERCBET 3 - a forecaster for epidemic development of *Cercospora beticola*. *OEPP/EPPO*(37), 344-349.

Racca, P., Jörg, E., Mittler, S., Petersen, J. CERCBET 3 - a forecaster for epidemic development of *Cercospora beticola*. . Brüssel,

Racca, P., Kleinhenz , B., Braun, C., Varrelmann, M. (2010a). ERYBET, UROBET & RAMUBET – die Prognosemodelle für die integrierte Bekämpfung der Zuckerrübenkrankheiten Mehltau, Rost und Ramularia. *57. Deutsche Pflanzenschutztagung. 6. - 9.September 2010 Humboldt-Universität zu Berlin; Gesunde Pflanze, gesunder Mensch*, D. Pflanzenschutztagung, J. Kühn-Institut, eds., Julius Kühn-Inst., Bundesforschungsinst. für Kulturpflanzen, Berlin, 156.

Racca, P., Kleinhenz, B., Jörg, E. (2007). SIMPEROTA 1/3 - a decision support system for blue mould disease of tobacco. *OEPP/EPPO*(37), 388-373.

Racca, P., Räder, T., Hau, B., Jörg, E. (2008). A decision support system for controlling leaf rust in winter rye and winter wheat. *Journal of Plant Pathology*, Torino, Italy, 507.

Racca, P., Tschöpe, B. (2010). SIMCOL1+3: Erarbeitung eines Entscheidungshilfesystems zur Optimierung der Bekämpfungsstrategie für die Anthraknose der Blauen Lupine (Poster). *57. Deutsche Pflanzenschutztagung. 6. - 9.September 2010 Humboldt-Universität zu Berlin; Gesunde Pflanze, gesunder Mensch*, D. Pflanzenschutztagung, J. Kühn-Institut, eds., Julius Kühn-Inst., Bundesforschungsinst. für Kulturpflanzen, Berlin, 431.

Racca, P., Zeuner, T., Jung, J., Kleinhenz, B. (2010b). Model Validation and Use of Geographic Information Systems in Crop Protection Warning Service. in *Precision Crop Protection - the Challenge and Use of Heterogeneity*, E.-C. Oerke, R. Gerhards, G. Menz, R. A. Sikora, eds., 259-276. Springer Netherlands, 978-90-481-9277-9.

Räder, T., Racca, P., Jörg, E., Hau, B. (2007). PUCREC/PUCTRI - a decision support system for the control of leaf rust of winter wheat and winter rye. *OEPP/EPPO*(37), 378-382.

Reynolds, M. R., Burkhart, H. E., Daniels, R. F. (1981). Procedures for statistical validation of stochastic simulation models. *For.Sci.*, 27, 349-364.

Röhrig, M., Sander, R. (2004). ISIP - online plant protection information in Germany. in *Online Agrometeorological Applications with Decision Support on the Farm Level. Cost Action 718: Meteorological Applications for Agriculture*, I. Thysen, A. Hočevar, eds., 41-47. Dina publication.

Roßberg, D., Gutsche, V., Kleinhenz, B. (2001). Prognose von Phytophthora infestans mit den SIMPHYT Modellen. *Gesunde Pflanzen*, 53(2), 37-43.

Roßberg, D., Jörg, E., Falke, K. (2005). SIMONTO - ein neues Ontogenesemodell für Wintergetreide und Winterraps. *Nachrichtenbl. Deut. Pflanzenschutzd.*, 57(4), 74-80,0027-7479.

Roßberg, D., Jörg, E., Kleinhenz, B. (1999). SIMLEP2 - ein Modell zur schlagspezifischen Prognose des Kartoffelkäferauftretens. *Nachrichtenbl. Deut. Pflanzenschutzd.*, 51(4), 81-87.

Roßberg, D., Racca, P., Jörg, E., Kleinhenz, B. (2000). Erste Erfahrungen mit dem Modell CERCBET1. *Nachrichtenbl. Deut. Pflanzenschutzd.*(52), 153-159.

Rossi, V., Battilani, P. (1986). Impiego di variabbili climatiche nella previsione delle infezioni primarie di Cercospora beticola Sacc. *Informatore Fitopatologico*(2), 29-38.

Rossi, V., Battilani, P. (1991). CERCOPRI: A forecasting model for primary infections of cercospora leaf spot of sugarbeet. *Bulletin OEPP-EPPO bulletin*(21), 527-531.

Rossi, V., Racca, P., Battilani, P. (1994). A simulation model for Cercospora leaf spot epidemics on sugarbeet. *Phytopath. Medit.*(33), 105-112.

Rossi, V., Racca, P., Giosuè, S., Battilani, P. (1997a). Decision support systems in crop protection: from analysis of the pathosystem to the computerized model. *Petria*, 7 Suppl.1, 7-26.

Rossi, V., Racca, P., Giosuè, S., Pancaldi, D., Alberti, I. (1997b). A simulation model for development of brown rust epidemics in winter wheat. *European Journal of Plant Pathology*, 103, 453-465.

Sargent, R. G. (1998). Verification and validation of simulation models. 121-130.

Schlesinger, S. (1979). Terminology for model credibility. *Simulation 34*, 3, 101-105.

Teng, P. S. (1981). Validation of computer models of plant disease epidemics: A review of philosophy and methodology. *Journal of Plant Diseases and Protection*, 88(1), 49-63,0340-8159.

Teng, P. S. (1985). A comparison of simulation approaches to epidemic modeling. *Ann. Rev. Phytopathol*, 23, 351-379.

Tschöpe, B., Racca, P. (2010). Die Ontogenese der Lupine: Modellierung und Validierung (Poster). *57. Deutsche Pflanzenschutztagung. 6. - 9.September 2010 Humboldt-Universität zu Berlin; Gesunde Pflanze, gesunder Mensch*, D. Pflanzenschutztagung, J. Kühn-Institut, eds., Julius Kühn-Inst., Bundesforschungsinst. für Kulturpflanzen, Berlin, 431.

Welch, S. M., Croft, B. A., Michels, M. F. (1981). Validation of past management models. *Environ. Entomol*, 10, 425-432.

Wernecke, P., Claus, S. (1996). Modelle der Ontogenese für die Kulturarten Winterweizen, Wintergerste und Winterraps. in *Reaktionsverhalten von agrarischen Ökosystemen homogener Areale*, H. Mühle, S. Claus, eds., 105-120. B.G. Teubner Verlagsgesellschaft, 3-8154-3529-3, Leipzig.

Zeuner, T. (2007). Landwirtschaftliche Schaderregerprognose mit Hilfe von Geographischen Informationssytemen.Dissertation, Johannes Gutenberg Universität, Mainz.

Zeuner, T., Kleinhenz, B. (2007). Use of geographic information systems in warning services for late blight. *OEPP/EPPO*(37), 327-334.

Zeuner, T., Kleinhenz, B. (2008). Optimierte Schaderregerprognose durch die Nutzung von Geographischen Informationssystemen (GIS). *56. Deutsche Pflanzenschutztagung in Kiel 22.-25. September 2008. Pflanzenproduktion in Wandel - Wandel im Pflanzenschutz?*, J. Kühn-Institut, P.-T. Deutsche, eds., Julius Kühn-Institut, Kiel, 266.

Zeuner, T., Kleinhenz, B. (2009). Verbesserte Schaderregerprognosen durch GIS Flächengenaue Modelle durch Nutzung von Geographischen Informationssystemen. *Kartoffelbau*, 4, 108-110.

17

Determination of Effective Policies for Ecological Agriculture Development with System Dynamics and Agent Based Models – Case Study in Slovenia

Črtomir Rozman, Andrej Škraba, Karmen Pažek,
Miroljub Kljajić, Martina Bavec and Franci Bavec
Univeristy of Maribor/Faculty of Agriculture and Life Sciences
Univeristy of Maribor/Faculty of Organizational Sciences
Slovenia

1. Introduction

Agricultural activity, beyond its primary function, can also shape the landscape, provide environmental benefits such as land conservation, sustainable management of renewable natural resources and preservation of biodiversity, and contribute to the socioeconomic viability of many rural areas (Majkovič et al., 2005). One way of emulating the prevailing EU policy reform trends is also to support and encourage organic farming, which is gaining in importance in Slovene agricultural production. Contemplated as a whole, any sound agricultural reform would entail not only necessary positive shifts in economic efficiency levels concerning the production and processing of food, but should specifically address some key socio-economic issues that are at the core of preserving and maintaining the ecological balances in the Slovene countryside; with biodiversity becoming an increasingly important agricultural policy concern (Ivančič et al., 2003). With respect to terms of multifunctionality, organic agriculture is the highest environmentally valuable agricultural system (Rozman et al., 2007a, 2007), and has strategic importance at national level that goes beyond the interests of agricultural sector.

This alternative agricultural paradigm may provide the link between objectives of sustainable resource use and sustainable regional development. The consequences of policies are long term and irreversible. In this light the conceptual methodological approach for evaluation of development policies for organic farming must be developed. Organic agriculture represents a complex system at national level (Shi and Gill, 2005) and different modeling approaches have been described in the literature (farm level, regional level and national level). Also, technologic economic simulation at farm level and multicriteria decision analysis are often used for decision support at farm level (Rozman et al, 2005; Pažek et al, 2006). Boorsma (1990) distinguishes three approaches in modelling the behaviour of the farmer: econometric modelling (based on linear regression equations of a data set); mathematical programming and modelling decision processes based on decision rules. At the national and regional level we often encounter econometric models that can

efficiently reflect the situation in agricultural systems and can also be used for forecasting policy consequences (Akinwumi et al., 2000, Turk (1998)). Although econometric models have a great methodological value and forecasting capabilities the modeler must ensure relatively long and consistent data series that are rarely available. Mathematical programming is frequently applied in farm planning. It allows determination of an optimal allocation of land, labour and capital, given a set of goals (e.g. maximisation of income and leisure and minimisation of risk) and constraints (e.g. labour and land). Boutkes and Van Keuler (2003) argue that the study of agricultural systems requires the use of non-linear dynamic models that allow simulation of the system in a qualitative way, based on a description of the underlying processes. Their approach is illustrated with a regional model that has been developed to simulate agricultural development in the Koutiala region in the south-western part of Mali. There are many factors, such as farm type and soil quality, that might influence farmers' decisions. However, attempting to consider the complex interactions of all factors in a single model is not a productive approach. Hence, the authors (Kaufmann et al., 2009) adopted the approach of isolating parts of a system and examining it assuming that all other things are equal. The diffusion of organic farming practices is modeled by a generic agent model (Borchev and Filippov, 2004) based on Theory of planned behavior for understanding and modeling the farmers decision making process.

System dynamics (SD) methodology (Forrester, 1961) can be and has been used as an alternative to econometric and mathematical programming approaches (Bockerman et al (2005)); Elshorbagy et al (2005)). SD model in its essence, is a continuous model because it is presented as a system of non-linear differential equations (Munitič and Trosić, 1997). There have been many important SD applications in the field of agriculture recentlyShen et al. (2009) present system dynamics model for the sustainable land use and urban development in Hong Kong. The model is used to test the outcomes of development policy scenarios and make forecasts. It consists of five sub-systems including population, economy, housing, transport and urban/developed land. Similar approach is presented by Weber et al. (1996).

However, the most important work in the field of simulation of development policy scenarios are presented by Shi and Gill (2005) who developed a system dynamics based simulation model for ecological agriculture development for Jinshan County (China) and Kljajić et al. (2000, 2001, 2002, 2003) with an integrated system dynamics model for development of Canary Islands where main interactions between agriculture, population, industry and ecology were taken into consideration. The preliminary results of SD simulation of organic farming development is conducted by Rozman et al. (2007) and Škraba et al. (2008). The model incorporates key variables affecting the organic farming systems and is used in identification in of main reasons that the strategic (15% or organic farms) has not been achieved. Yet this research did not incorporate the full aspects of food market and consumer factor (Rozman et al., 2007). However, consumer concerns are inherently dynamic because they respond to difficult and complex societal and technological situations and developments. For example, because of the rising concern with global warming, carbon dioxide absorption of crops is now attracting public attention, which means that new requirements are being proposed for the environmentally friendly production of crops (Korthals, 2008). In this light Rozman et al. (2008) and Rozman et al. (2010) upgraded the model with the inclusion of organic market development factor.

This paper presents a system dynamics model for development of organic agriculture in Slovenia in order to identify key reasons and propose development policy to achieve strategic goals set in the ANEK (Majcen and Jurcan, 2006). The paper is organized as follows: first we present the state of the art of organic agriculture with its system analysis and identify key variables, main flows and feedback loops in the systems. The results section presents scenarios (different policies in organic farming) and their evaluation using the developed SD model. Main findings and suggestions for further study conclude this article.

2. Model development

2.1 Study area

We selected the Republic Slovenia as study area in order to develop and employ the SD model. The most of Slovenia agriculture is located (with exception of eastern flat land with its intensive field crop production) in hilly unfavorable areas. In the European space Slovenia belongs to the countries with the most unfavorable conditions because of its diverse and mountainous relief and high proportion of carst areas. Recent studies have also shown deficiencies of organic products on the market (Pažek at al., 2006). Thus organic agriculture has been identified as one of developmental opportunities.

There are approximately 80,000 farms in Slovenia; conventional and organic. In year 2006 only 1,728 farms are in the organic farm control system. Even though the subsidy has been offered (Recent research has shown (Rozman et al., 2007) that correlation between subsidies level and number of organic is too low) to the farmers, the proportion of the organic farms is still low, not higher than 5%. The short term strategic goal is to reach the 10% or 15% ratio by the year 2015. This is determined by the state action plan ANEK (Majcen and Jurcan, 2006). Although the number has increased to 2000 in 2007 and 2067 in 2008 the strategic goal (15%) is still underachieved.

In Slovenia up to 440,349 hectares are defined as less favoured areas (LFA). These are hilly and mountainous areas, areas with karst features or other factors that limit possibilities of intensive farming. Relatively high share of less favourable areas make Slovenia suitable for less intensive sustainable production systems – such as organic agriculture.

The system analysis of organic agriculture

In order to provide the proper systemic solution of the described problem, the simulation model should be build which represents the structure with key elements. The simulation model should consider the key variables that influence the development of the organic farming such as:

- number of conventional farms
- number of organic farms
- conversion
- subsidies
- promotion of organic farming (marketing, market development, education
- organization of general organic farming support environment
- system self awareness
- delay constants of process changes

The key variable in the model is the number of organic farms. These are the farms that are in the control system at the one of the control organizations. The growth of the number of organic farms was initially (in year 1998) almost linear however, in the years from 2003-2005 the growth is moderated to approximately 4% despite the increase of subsidies for 20%-30%.

Fig. 1. Map of Republic of Slovenia (relief)

At the development of the causal loop diagram (Fig. 2.) as the first step of the development of SD model the following key variables were identified:

1. the number of potential candidates (farms) for conversion to organic farming
2. the number of farms converted to organic farming
3. the flow between (1) and (2): conversion rate (transition)

Loop Ⓐ represent negative loop with the goal value of 0 (depleting the number of "Conventional Farms"). Number of "Conventional Farms" divided by the "Total Number of Farms" yields the "Concentration of Conventional Farms" which is initially high meaning that there should be high initial preference for "Conversion". "Concentration of Conventional Farms" therefore positively influences the "Communication". This variable represents the general communication between the conventional approach members and organic approach members. "Conversion" positively influences on the number of "Organic Farms". If the number of "Organic Farms" increases, the "Information Spread" increases above the level that would otherwise have been. "Information Spread" by "Organic Farms" member is positively influenced by the "Information Spread Factor" which could be for

example increased by the marketing campaigns. "Information Spread" positively influences on the "Communication". The number of "Conversion" is determined by the "Success Factor" which determines the "Communication Success" yielding the number of convinced conventional members that decide to make a "Conversion". Loop Ⓑ is reinforcing feedback loop compensated by the initial balancing feedback loop marked with Ⓐ. If the number of "Organic Farms" increases, the "Promotion and Marked Development" supported by the "Policy Support Factor" increases above the level that would otherwise have been. Higher "Promotion and Market Development" positively influences the "Self Organization Resources", which positively contribute to the "Support Resources" on which the "Conversion" is dependent on.

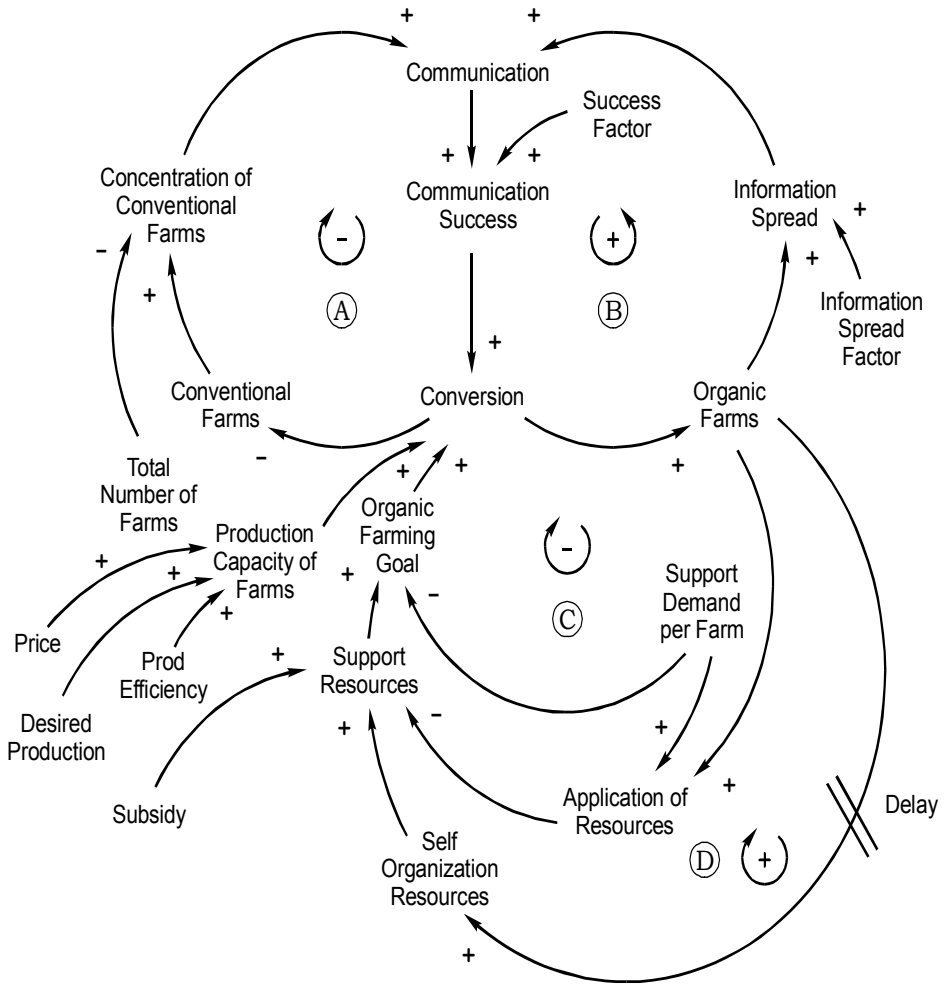


Fig. 2. Causal Loop Diagram (CLD) of system structure

There is the delay mark between the "Promotion and Market Development" and "Self Organization Resources". Here the longer delays should be considered since there is a significant amount of time needed in order to promote the organic farming idea and

marketing channels which would support the organic farming. "Support Resources" are significantly depended on the government "Subsidy". More there are "Support Resources" higher the "Organic Farming Goal" is set meaning, that larger number of organic farms could be supported. If the "Organic Farming Goal" increases, the "Conversion" increases above the level that would otherwise have been.

Mentioned interconnections marked with Ⓓ has a characteristic of reinforcing feedback loop. By proper government policy the growth in the number of "Organic Farms" should be properly supported in order to promote increase in self organization of e.g. organic food marketing and promotion.

Therefore the reinforcing feedback loop Ⓓ should be applied as the growth generator in the system.

Loop Ⓒ represent balancing loop. If the number of "Organic Farms" increases, the "Application of Resources" increases above the level that would otherwise have been. "Application of Resources" is also dependant on the resources needed per farm i.e. "Support Demand per Farm". Higher "Application of Resources" cause the depletion of the "Support Resources". "Organic Farming Goal" is dependant on the "Support Demand per Farm". If there is more resources needed per farm less organic farms could be supported therefore lower number of "Conversion" could be expected in such case. In considered real case, the negative loops Ⓐ and Ⓒ are dominant leaving the system in unwanted equilibrium state. This would mean, that the number of organic farms is constant well below desired. In order to move the system from the equilibrium one should consider the policies which would raise the impact of reinforcing feedback loops Ⓑ and Ⓓ which should move the system state i.e. number of "Organic Farms" to the higher equilibrium values. Price, Desired production and Production efficiency are also important factors that impact the transition intensity.

Model development

System dynamics model structure is shown in figure 3. Model consists of 29 variables and 51 links.

There are two level elements applied in the upper part of the model. The variable "conventional_farms" represent the number of conventional farms. By the flow "transition" the "conventional_farms" become "organic_farms". This structure is commonly known as the market absorption model. "conversion" is dependent on the "organic_farming_goal". The goal is set by the "support_resources" available modeled as a level element. The conversion could only be achieved if there is enough "support_resoureces" present in order to make a "conversion". The "support_resoures" are not only the financial means. Here the society support is also considered, for example education should provide positive thinking when organic farming is considered. In this category the market development as well as the demand should also be considered. However at present the "support_resources" are mainly dependent on the subsidies form the government. Important variable "self_organization_resources" is driven by the impact of the policy and society support which intensifies with the number of "organic_farms". This represents the application of reinforcing feedback loop which should be augmented. "development_limit" represents the function which consider variable consumption of the resources. If the resources are scarce the usage is lower than in the case of abundance. Resources are consumed by the

"organic_farms". The prosperity of the "organic farms" therefore depends on the "support_resources" which are not only financial means; here the social impact of organic farming represents the supportive environment which should sustain such an activity which is in the world of consumption counterintuitive (Forrester, 1961).

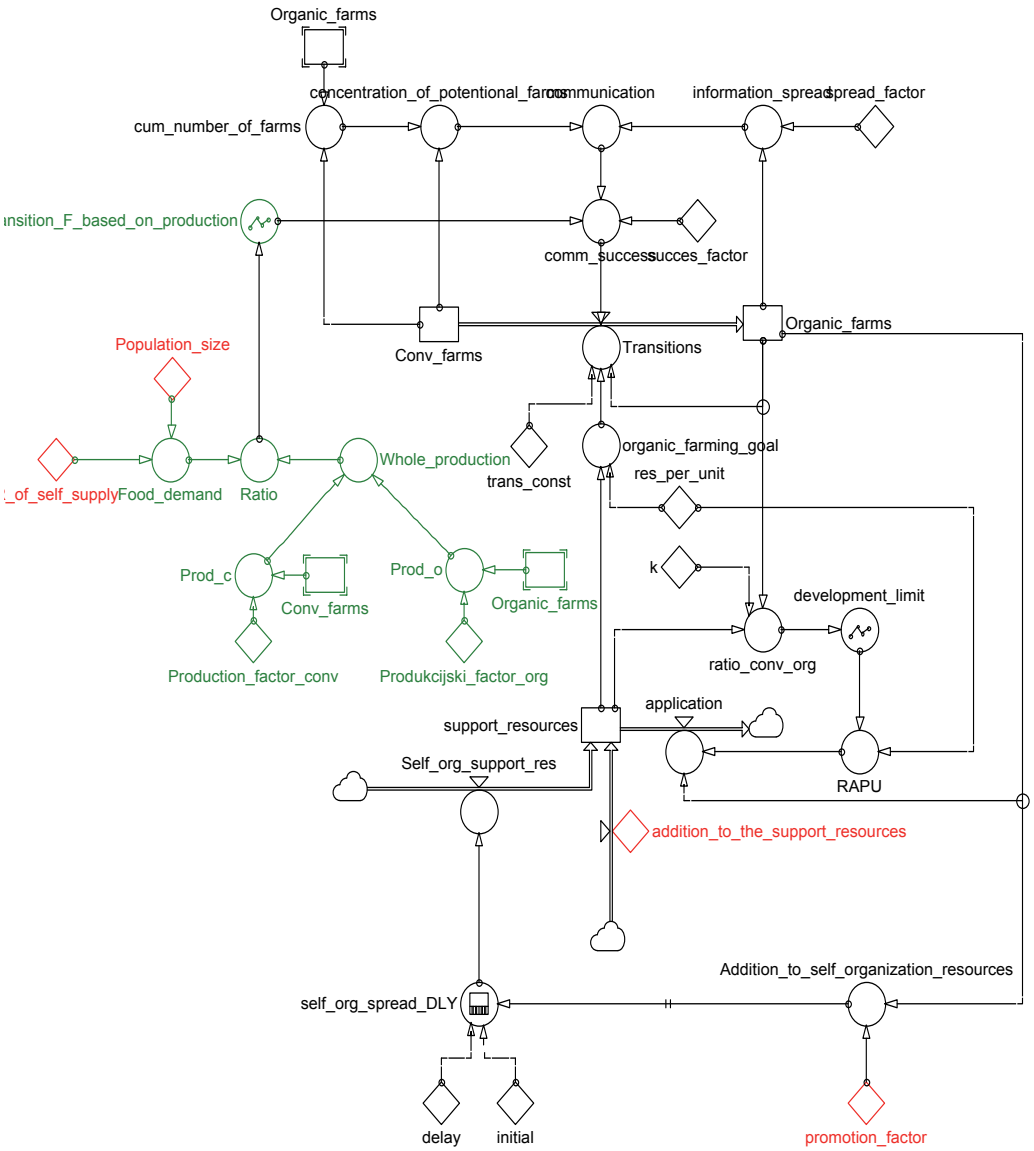


Fig. 3. System Dynamics model of Organic Farming Development

Figure 4 shows model examples of model equations. There are 77,000 conventional farms initially and 1,728 organic farms. The model is realized in Powersim. By the following equations the model could easily be transformed to other SD tools such as Vensim, iThink, Stella etc.

```
INIT   CONV_FARMS = 77000
FLOW   CONV_FARMS = -DT*TRANSITIONS
DOC    CONV_FARMS = THE NUMBER OF CONVENTIONAL FARMS WHICH ARE
CANDIDATE FOR ENTERING THE CONTROL SYSTEM.
UNIT   CONV_FARMS = FARM
INIT   ORGANIC_FARMS = 1728
FLOW   ORGANIC_FARMS = +DT*TRANSITIONS
DOC    ORGANIC_FARMS = ORGANIC FARMS.
UNIT   ORGANIC_FARMS = FARM
INIT   SUPPORT_RESOURCES = 500
FLOW   SUPPORT_RESOURCES = +DT*SELF_ORG_SUPPORT_RES
       +DT*ADDITION_TO_THE_SUPPORT_RESOURCES
       -DT*APPLICATION
DOC    SUPPORT_RESOURCES = SIZE OF SUPPORT RESOURCES DETERMINES
THE ATRACTIVENESS OF ORGANIC FOOD PRODUCTION.
UNIT   SUPPORT_RESOURCES = FARM
AUX    APPLICATION = RAPU*ORGANIC_FARMS
DOC    APPLICATION = APPLICATION OF SUPPORT RESORUCES DIMINISHES
THE AVAILABILITY. CONSEQUENCE: LOST OF FOCUSS AND ACTRACTIVENESS.
UNIT   APPLICATION = FARM/YEAR
AUX    SELF_ORG_SUPPORT_RES = 1000+SELF_ORG_SPREAD_DLY
DOC    SELF_ORG_SUPPORT_RES = HOW MANY FARMS COULD BE SUPPORTED BY
SELF-SUPPORT ACTIONS.
UNIT   SELF_ORG_SUPPORT_RES = FARM/YEAR
AUX    TRANSITIONS            =            ((ORGANIC_FARMING_GOAL-
ORGANIC_FARMS)/TRANS_CONST)*COMM_SUCCESS
DOC    TRANSITIONS = TRANSITION FROM CONVENTIONAL FARMS TO ORGANIC
FARMS.
UNIT   TRANSITIONS = FARM/YEAR
AUX    ADDITION_TO_SELF_ORGANIZATION_RESOURCES                    =
PROMOTION_FACTOR*ORGANIC_FARMS
DOC    ADDITION_TO_SELF_ORGANIZATION_RESOURCES = EFFORT TO PROMOTE
SELF-ORGANIZING RESOURCES.
UNIT   ADDITION_TO_SELF_ORGANIZATION_RESOURCES = FARM
AUX    COMM_SUCCESS                                               =
SUCCES_FACTOR*COMMUNICATION*TRANSITION_F_BASED_ON_PRODUCTION
DOC    COMM_SUCCESS = CONVICED FARMERS DECIDE TO TRANSFORM.
UNIT   COMM_SUCCESS = FARM/YEAR
AUX    COMMUNICATION =
```

Fig. 4. Examples of model equations

Agent-based approach

In our research the agent-based approach has been considered as the possible way to analyze the dynamics of transition to organic farming. By this, one could compare both methodologies, System Dynamics and Agent-Based modeling. In Agent based model built with AnyLogic (Borschev and Filippov, 2004) shown in figure 4; we define the agents as farms. The model is represented by two agent states; 1) Conventional Farms (red) and 2) Organic farms (green). Transition among particular states is determined by the promotion of organic farming and information spread. The contacts in the state of organic farming is also

considered. This approach is promising since it is possible to model whole agricultural sector where each particular farm is taken into account. Initially one initializes the particular number of agents, in our case 2000, since this is the number of potential farms for transition. The model is based on the Bass diffusion agent-based model. The number of farms is set to 2000 since the agent model with 20.000 farms would take too much time to run. Initially all the agents are painted red since all the farms are conventional. During the simulation agents transform from conventional to organic farms, which could be observed on the graphical view; the agent turns from red to green. Since the agents could transform from conventional farms to ecological in two ways there are two different border representations. If the agent performs transition on account of the promotion, the border of the agent turns yellow. If agent performs transition on account of other causes, the border turns blue. In this manner one could easily estimate how many agents performed transition in particular was as well as how fast particular transition occurred during simulation.

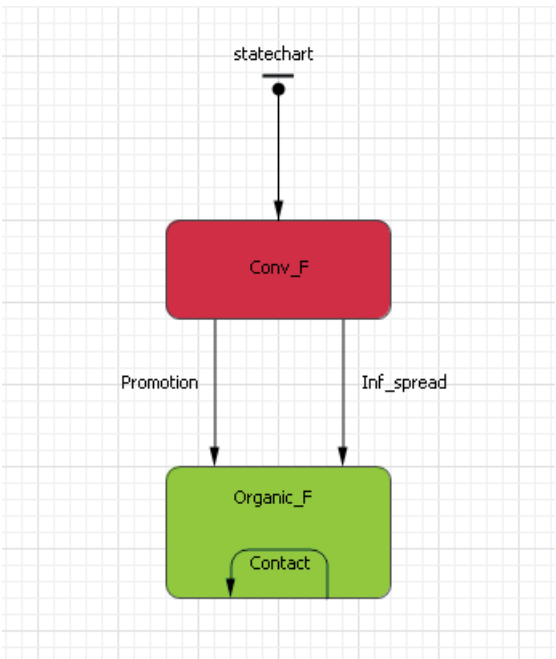


Fig. 5. Agent-based simulation model of transition to organic farming

3. Results

Table 1 shows the parameter values for the eight scenarios performed on the developed system dynamics model. SC1 is the initial scenario where the initial amount of the subsidy is provided (1000). This would mean that there are some resources provided by subsidy to support 1000 organic farms.

Figure 5 shows results of eight different simulation scenarios. One of more important findings is, that the system is sensitive to the changes in demand. If one observe scenario 7 and 8, where the population is changed only for 50k and 100k, once could observe, that the conversion to the organic farming would be jeopardized.

Scenario	Subsidies	Self-supply coefficient	Delay	Promotion factor	Population
1	1000	1,3	1	0,8	2M
2	3000	1,3	1	0,8	2M
3	1000	1,2	1	0,8	2M
4	2000	1,2	1	0,8	2M
5	2000	1,2	36	0,8	2M
6	2000	1,2	12	1	2M
7	2000	1,2	12	1	2.05M
8	2000	1,2	12	1	2.1M

Table 1. Values for particular scenarios considering amount of subsidies, self-supply coefficient, delay which represents to what proportion self-supply of organic farms should be considered.

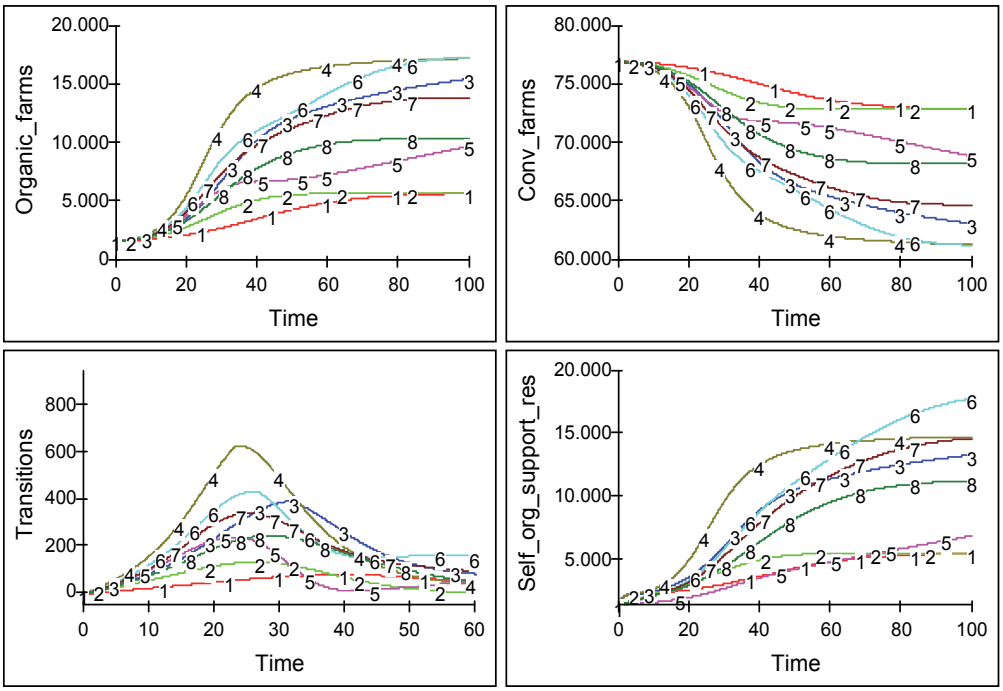


Fig. 5. Example of seven performed scenarios showing dynamics of Eco farms, Conventional farms, Transitions and Self-organizing support resources

As the mean of concept validation the results of agent-based model are shown in the following section. Tab. 2. shows list of parameter values for Agent Based Model for four different scenarios which are performed as the demonstration of how future Agent Based Model should be implemented.

Figure 6 show the results of first simulation scenario SCA1. At the beginning the transition is started with low gradient until, on account of promotion, the gradient increases as well as

number of agents. Informal information contributes to more intensive conversion until the proportion of conventional farms is low and informal communication loses its power. As one could observe, the farms that are unchanged are on the outskirt of the system due to remoteness and lover intensity of communication with other farms. Such farms are consequently not given the same amount of informal promotion.

Scenario	No. of farms	Effect of support	Transition factor	No. of contact
SCA1	2000	0.011	0.015	125
SCA2	2000	0.1	0.005	3
SCA3	2000	0.05	0.025	180
SCA4	2000	0.01	0.015	300

Table 2. List of parameter values for Agent Based Model; four different scenarios

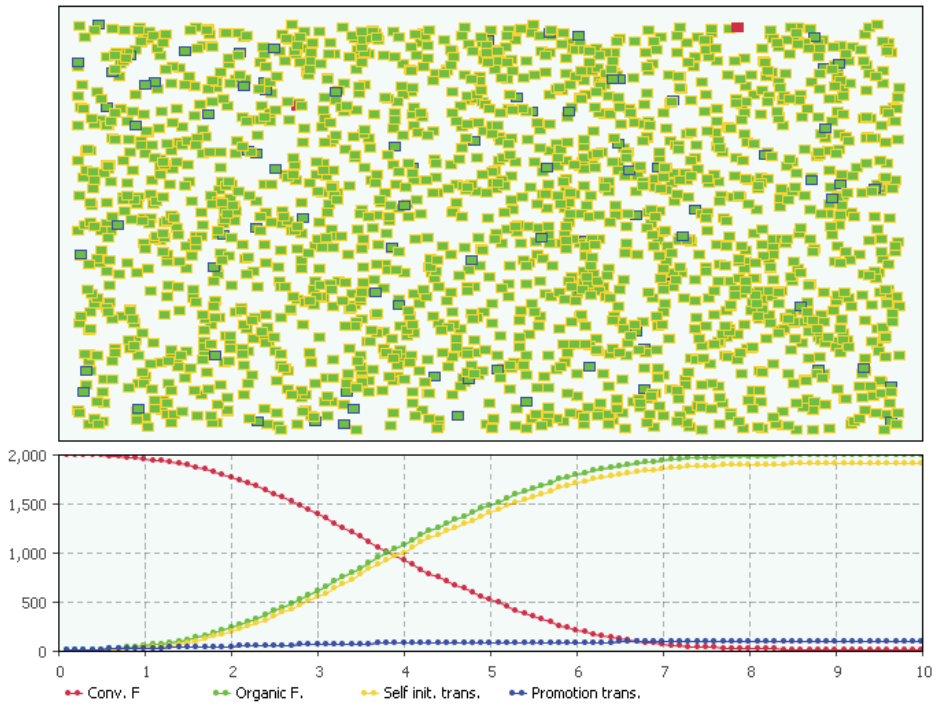


Fig. 6. Results of SC1, ABM

Figure 7 shows second scenario SCA2, where informal information flows are considered, here it is put to minimum. Increase of the model is almost linear; here the promotion of organic farming dominates in its influence. Here the importance of informal communication could be observed as well as impact of certain promotion actions, which is tied to particular number of organic farms. The proper influence of promotion is also confirmed. In the real world, such situation would occur in the case of very isolated farms, which have no proper contacts with other farms and limited access to support resources despite the fact, that the support level might be high.

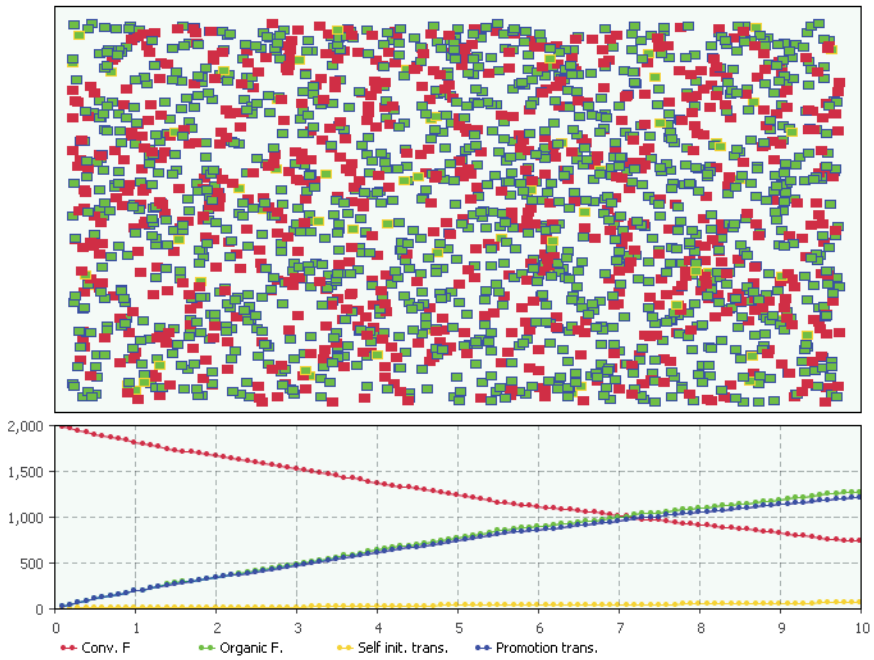


Fig. 7. Results of SC2, ABM

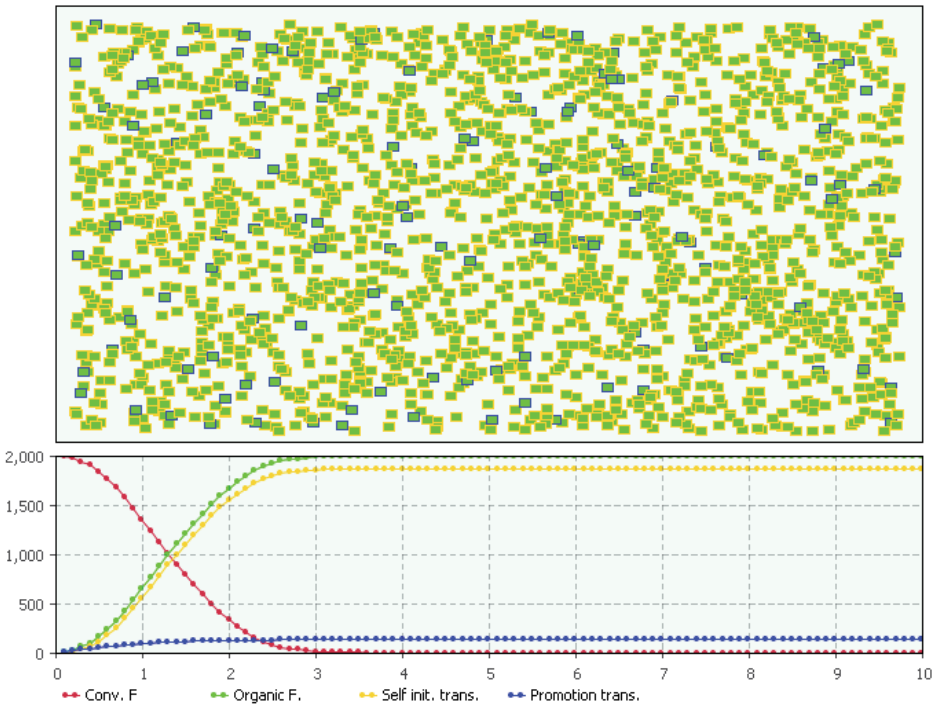


Fig. 8. Results of SC3, ABM

Figure 8 shows third scenario SCA3, here the successfulness of promotion is lowered to minimum. On the contrary, the level of communication and transition intensity is increased as the consequence of communication i.e. promotion. As one could observe, after the initial starting time, significant increase in transition occur. One could conclude, that promotion »infects« few initial agents, which, due to the high level of communication contribute to the explosion of transitions. In the real situation this would mean, that the agents have low level of susceptibility for promotion however, they are strongly interconnected and demonstrate large level of interpersonal trust.

Figure 9 shows the fourth simulation scenario SCA4, where the key role is played by the communication between agents. The level of the communication is increased on the highest value. The parameters of promotion and the intensity of transition are lower than in the third scenario. However, the transition is exceptionally fast. One could observe, that several agents become isolated, those, who have less intensive contacts. In the real world, such situation would occur if the cooperation among agents (farms) would be very strong with strong contacts. This could be achieved via internet and other means of e—communication, personal contacts etc. Here the technology as well as support action in the field of communication should be considered.

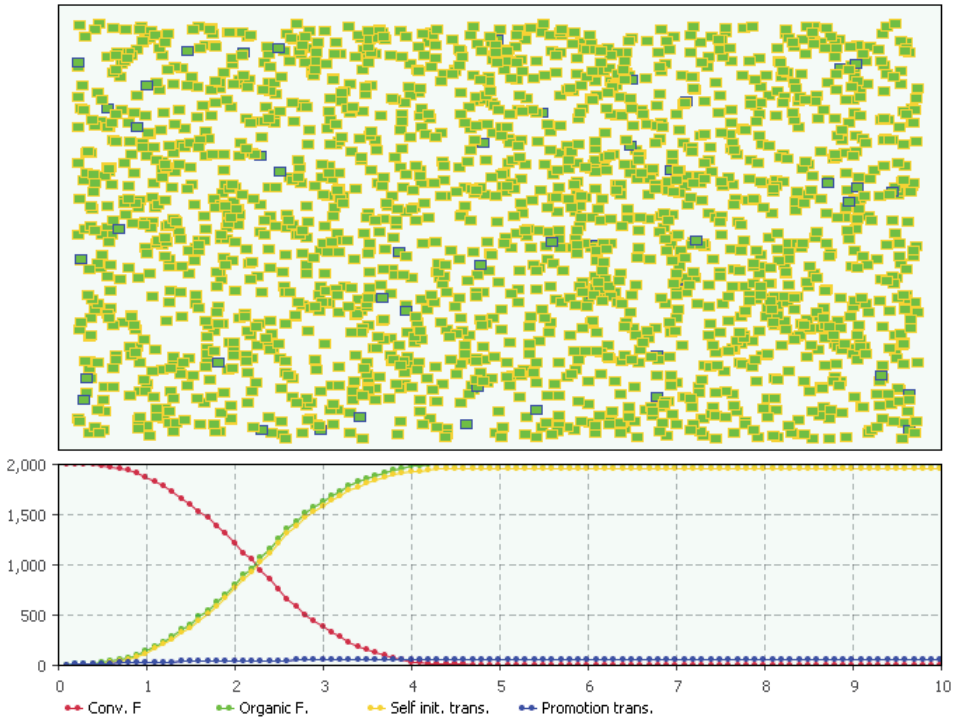


Fig. 9. Results of SC4, ABM

4. Discussion

Promotion factor represents the policy to promote organic farming and self-organizational resources. That would mean the development of organic-farming marketing, production

etc., which would contribute to better demand. This value is set to 0.8 initially which means, that each new organic farm rises the resources (not only financial) for 0.8 additional organic farm by adding, e.g. to the better development of the organic marketing and prodution. The delay represents the number of months in order to spread the effect of the additional support resources in the system. Initially we consider, that this delay is short, in our case 1 month. "1" marks the response of first scenario, SC1. Self-supply coefficient represent the proportion to which the country should be self content regarding the food supply. This factor determines the food demand. 1.3 would mean, that the desired food production shoud be 30% higher of normal production. The coefficient of self-supply determines the demand which also depend on the Whole production of the agriculutral sector. Here it is important, that in the case of higher prices, the food production capacity would play a key role and influence the possible negative transitions (back to conventional farming). Population is considered as 2 million which determines the food demand. If one compares it to the scenario "2" where the subsidies are rised to 3000 the more intensive transitions are observed. However, the observed number of organic farms is far from desired meaning, that the subsidies by themselves would probabbly not be enough. In the scenario SC3, the impact of decreasing self-supply coefficient is considered as well as decrease in subsidies. If one decreases Self-supply coefficient, the demand/supply delay ratio would be better, influencing on the better demend for organic farming products. This would in turn compensate the lower Subsidies and provide the highest conversion so far. In scenario SC4 the subsidies are rised to 2000 which gives the best results regarding the response of the system and the limit value of the organic farms, which is approximately 17,000. This would mean, that the right political choice would be, to increase demand for the organic farming products by lowering the self-supply and provide larger amount of subsidies. However, this could be risky in the condition of higher food prices. SC5 considers higher delay at the establishement of the self-supporting resources, which is set to 36 months. This is more realistic since the establishment of self-supporting resources takes some time. The consequence is, that the rise in the number of farms is much slower. This would mean, that it is very important, to quickly establish self-support resources for organic farming if we want to achieve fast transitions. SC6 shows the impact of lowering the delay in establishing self-support resources. Here the delay is put to the 12 months giving much better response and achieving the limit value of the organic farms, which is approximately 17,000. SC7 shows the impact of larger food demand in the case that the population would increase. This would have for the consequence larger food demand and rise in prices. It such situation, the transition would be slower and less farmers would choose to switch to the organic farming due to higher food prices. SC8 shows even worse situation if the population would have an additional increase meaning, that the demand for food would be even higher. In that case, the transition to organic farming would be even slower. One of the important questions is »How could the subsidies be replaced?« As the model shows, the main leverage is the organic farming promotion and market development. In this manner, the self-supporting resources are established which further promote the transition to the organic farming. This is the counterpart of direct subsidies which should be converted to the actions that support self organization component in the system. The presented combined methodological framework (SD) for the analysis of development of organic farming could provide additional information support to agricultural policy makers, bring additional clarity to the decision, and could therefore play an important role in further development of organic farming, in particular as assistance and advisory in policy planning.

5. Conclusion

In this paper an attempt was made to employ system dynamics model in order to simulate the development of organic agriculture. The presented SD model enables simulation of different policies and this kind of model is comprehensible to a wide range of users in the decision making process.

After performing several simulation scenarios the following findings could be abstracted:

- Conversion to the organic farming relies on subsidies which provide the main source of conversion from conventional farming to organic farming.
- Subsidies are not the only driving force in the system; even more important are other activities that promote organic farming.
- Subsidies could not be provided in sufficient amount in order to complete conversion from the conventional to organic farming.
- Feasible strategy to achieve complete conversion should consider reinforcing feedback loop between resources, number of organic farms and supportive actions which are bounded to the number of organic farms.
- Current output parameter i.e. number of organic farms, is caught in an unwanted equilibrium value due to the domination of balancing feedback loops in the system.
- Important factor is self-organization of the organic farming environment which includes market development and general public awareness.
- Due to the large systemic delays in the system the anticipative value of the system control plays an important part.
- Important factor that influences the transitions to the organic farming is demand on the market which is largely driven by the politics and the self-supply principle.
- The agent based model shows that it is possible to build an agent-based model which would enable to monitor each particular farm and its transition. The tool AnyLogic has been identified as the proper tool for such modeling task.

Further strategic actions should consider the dynamic response of the system and the feasibility of stated system target values. Consideration of the interaction of four main feedback loops indicated in the system which determines the system performance provides the means for proper control strategy definition.

The presented combined methodological framework (SD) for the analysis of development of organic farming could provide additional information support to agricultural policy makers, bring additional clarity to the decision, and could therefore play an important role in further development of organic farming, in particular as assistance and advisory in policy planning. Further research is needed in the field of SD modeling in order to properly evaluate the applicability of the proposed model. Especially the market development of organic food should be additionally considered as proposed by Rozman et al., (2008). The SD model should be further verified and correspondingly improved. The agent-based model should be developed which would enable precise monitoring of each particular farm. The model structure and its results should be evaluated by relevant expert group.

6. References

Akinwumi, A.A; Mbila, D.; Nkamleub, G.B. & Endamana, D. (2000). Econometric analysis of the determinants of adoption of alley farming by farmers in the forest zone of southwest Cameroon Agriculture, *Ecosystems and Environment*, Vol. 80, pp. 255-265

Bockermann, A.; Meyer, B.; Omannc, I. & Spangenberg, J.H. (2005). Modelling sustainability Comparing an econometric (PANTA RHEI) and a systems dynamics model (SuE), *Journal of Policy Modeling*, Vol. 27, pp. 189-210

Boorsma, A. (1990). *Possibilities for Arable Farms in the Dutch Peat Soil District.* PhD thesis, Rijks Universiteit Groningen, Groningen (Dutch).

Bontkes, T.S. & Van Keulen, H. (2003). *Modelling the dynamics of agricultural development at farm and regional level*, Agricultural Systems, Vol. 76, No.1, pp. 379-396

Dubois D., Resconi G. (1992); Hyperincurisvity - a new mathematical theory, 1st ed. Presses Universitaires de Liège.

Borschev, A. & Filippov, A. (2004). *From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools*. The 22nd International Conference of the System Dynamics Society, July 25 - 29, 2004, Oxford, England.

Elshorbagy, A.; Jutla, A. & Barbour, L. (2005). System dynamics approach to assess the sustainability of reclamation of disturbed watersheds, *Can. J. Civil Eng.,* Vol. 32, No.1, pp. 144-158

Forrester, J.W. (1961). *Industrial Dynamics*, MIT Press, Cambridge, MA, pp. 464

Hong S.H. (2008). Forest landscape models: Definitions, characterization, and classification, *Forest Ecology and Management,* Vol. 254, No. 3 (http://www.sciencedirect.com/science/article/B6T6X-4R16JMP-3/2/44b9afbe4381a00ac6d7616d18001896)

Ivančič, A.; Turk, J.; Rozman, Č. & Šiško, M. (2003). Agriculture in the Slovenia transition economy: The preservation of genetic diversity of plants and ethical consequences, *Journal of Agricultural and Environmental Ethics*, Vol. 16, pp. 337–365

Kaufmann P.; Stagl, S. & Franks, D.W. (2009). Simulating the diffusion of organic farming practices in two New EU Member States, *Ecological Economics*, Vol. 68, pp. 2580–2593

Kljajić M. (1998). Modeling and Understanding the Complex System Within Cybernetics. Ramaekers M. J. (Ed.), 15th International Congress on Cybernetics. Association International de Cybernetique. Namur, pp. 864-869

Kljajić, M.; Bernik, I.; & Škraba, A. (2000). Simulation Approach to Decision Assessment in Enterprises, *Simulation*, Vol. 75, No. 4, pp. 199-210

Kljajić, M. (2001). Contribution to the meaning and understanding of anticipatory systems. V: DUBOIS, Daniel M. (ur.). Computing anticipatory systems, (AIP conference proceedings, 573). Melville (New York): American Institute of Physics, 2001, pp. 400-411

Kljajić, M.; Cav, L. & Škraba, A. (2002). System Dynamics Model Development of The Canary Islands for Supporting Strategic Public Decisions. Proc. of the 20th International Conference of the System Dynamics Society; The System Dynamics Society, Palermo, Italy, pp.16

Kljajić, M.; Cav, L.; Škraba, A. & Peternel, J. (2003). Simulation Model of the Canary Islands for Public Decision Support – Preliminary Results. Proc. of the 20 th International Conference of the System Dynamics Society; The System Dynamics Society, Albany, NY

Kljajić, M.; Škraba, A.; Kofjač, D. & Bren, M. (2005). Discrete Cobweb Model Dynamics with Anticipative Structure, *WSEAS Transactions on Systems*, Vol. 4, No. 5

Korthals, M. (2008). Ethical Rooms for Maneuver and Their Prospects *Vis-à-vis* the Current Ethical Food Policies in Europe, *Journal of Agricultural and Environmental Ethics,* Vol. 21, No. 3

Majcen M.H. & Jurcan S. (Eds.) (2006). Action Organic Farming Development Plan in Slovenia to Year 2015 (ANEK), Government of Republic of Slovenia.

Majkovič, D.; Borec, A.; Rozman, Č.; Turk, J. & Pažek, K. (2005). Multifunctional concept of agriculture: Just an idea or the real case scenario? *Društvena istraživanja,* Vol. 14, No. 3, pp. 579–596

Munitič, A. & Trosić, Z. (1997). System dynamics continuous modelling of the ecological subsystem of the 'Kastela Bay', *Simulation Practice and Theory*, Vol. 5, No. 7-8, Selected papers of the 5th European Simulation Congress, 15 October 1997, pp. 701-717, ISSN 0928-4869, DOI: 10.1016/S0928-4869(97)00017-7. (http://www.sciencedirect.com/science/article/B6V19-3SP2RCD-9/2/f2a7df3b95f6ac8c2f6b4a62edf0c018)

Pažek, K.; Rozman, Č.; Borec, A., Turk, J., Majkovič, D., Bavec, M. & Bavec, F. (2006). The use of multi criteria models for decision support on organic farms. *Biol. agric. hortic.*, Vol. 24, no.1, pp. 73-89.

Rozman, Č.; Škraba, A.; Kljajić, M.; Pažek, K.; Bavec, M. & Bavec, F. (2008). The Development of organic agriculture model – a system dynamics approach to support decision making processes. In *Symposium on Engineering and Management of IT-based organizational system* (pp. 39-44). Baden-Baden, Germany: The international institute for advances studies in system research and cybernetics.

Rozman, Č.; Škraba, A.; Kljajić, M.; Pažek, K.; Bavec, M. & Bavec, F. (2007a). The system dynamics model for development of organic agriculture. In *D. M. Dubois* (Ed.), *Eighth International Conference on Computing Anticipatory Systems, HEC-ULg.* Liege, Belgium: *CASYS'07, in press.*

Rozman, Č.; Škraba, A.; Pažek, K. & Kljajić, M. (2010). *The System dynamics model for strategic planning of organic farming development*. V: KLJAJIĆ, Miroljub (ur.), LASKER, George Eric (ur.). Advances in simulation-based decision support : [papers presented at the 22nd International Conference on Systems Research, Informatics and Cybernetics held August 2-6, 2010 in Baden-Baden, Germany]. Tecumseln (Canada): The International Institute for Advanced Studies in Systems Research and Cybernetics, pp. 31-35.

Rozman, Č. & Pažek, K. (2005). Application of computer supported multi-criteria decision models in agriculture. *Agric. conspec. sci.,* Vol. 70, No. 4, pp. 127-134.

Rozman, Č.; Škraba, A.; Kljajić, M.; Pažek, K.; Bavec, M. & Bavec, F. (2007). The system dynamics model for development of organic agriculture. In: DUBOIS, Daniel M. (ur.). Eighth International Conference on Computing Anticipatory Systems, HEC-ULg, Liege, Belgium, *CASYS'07 : abstract book*, (Casys).

Shi, T. & Gill, R. (2005). Developing effective policies for the sustainable development of ecological agriculture in China: the case study of Jinshan County with a systems dynamics model, *Ecological Economics,* Vol. 53, No. 2, pp. 23-246

Shen, Q.; Chen, Q.; Tang, Bo-sin.; Yeung, S.; Hu, Y. & Cheung, G. (2009). A system dynamics model for the sustainable land use planning and development, *Habitat International*, Vol. 33, No. 1, pp. 15-25.

Šraba, A.; Rozman, Č., Klajić, M.; Pažek, K., Bavec, M., Bavec, F. & Kofljač, D. (2008). *Strategy for organic farming model development.* V: Dangefrield, B.C. (ed.). Proceedings of the 26th International Conference [of] the System Dynamics Society, Athens, Greece, July 20 - July 24, 2008. Albany (New York): The System Dynamics Society, cop, pp. 104-105

Weber, G.; Smith, J.; & Manyong, M.V. (1996). System dynamics and the definition of research domains for the northern Guinea savanna of West Africa, Agriculture, *Ecosystems & Environment*, Vol. 57, No. 2-3, pp. 133-148

18

Web-Based Decision Support for Sustainable Pest Management in Fruit Orchards: Development of the Swiss System SOPRA

Jörg Samietz, Benno Graf, Heinrich Höhn, Lukas Schaub
Hans Ulrich Höpli and Elisabeth Razavi
Agroscope Changins-Wädenswil Research Station ACW,
Wädenswil
Switzerland

1. Introduction

Plant protection in modern orchard management relies on precise timing of monitoring and control of pest populations. Particular stages within the life cycle of insects and other pests have to be surveyed throughout the season by different monitoring procedures in order to establish economic injury levels and to test for the need of control measures, respectively (Kogan, 1998). Growers need precise information about when these stages are present and corresponding techniques have to be applied or monitoring has to be carried out (Fig. 1). Especially in sustainable production applying integrated pest management (IPM) the pests necessarily have to be monitored in order to assess the risk and consequently to judge the requirement of an intervention with a control measure (Norton & Mumford, 1993; Dent, 1995; Pedigo & Rice, 2005). Precise timing of monitoring measures assures reliability of the results and saves time during the decision finding process (Fig. 1).

In case control is needed, all regarding actions taken need to be scheduled in relation to crop and pest phenology. Modern control measures even rely on more precise timing especially when modes of action aim at very specific developmental stages of the pest (Blommers, 1994). Besides increasing efficacy, the precise timing of plant protection measures reduces their side effects and may substantially reduce the number of treatments and thereby resources and money spent during the process.

The required knowledge on the phenology of the pest populations can be established by forecasting systems that at the best are connected with information on the pests and possible management options to decision support systems. Hitherto, temperature sums and more recently simulation models have been used in tree fruit growing to predict the phenology of pests and hence to facilitate timing of monitoring and plant protection measures.

However, aside from a few exceptions (e.g. Welch, et al. 1978; Morgan & Solomon, 1996), the simulation models to predict the phenology of fruit pests are often not designed to be used by growers, consultants or extension services. Also they are often based on very different approaches and programming languages or require special driving variables, which makes them even difficult to use by extension services (Rossing et al., 1999; van der Werf et al., 1999).

Here we introduce the forecasting tool SOPRA which has been developed to optimize timing of monitoring, management and control measures of major insect pests in Swiss fruit orchards. The system consists of a locally based user interface with the different species models and a web-interface to provide simulation results and decision support to consultants and growers (www.sopra.info).

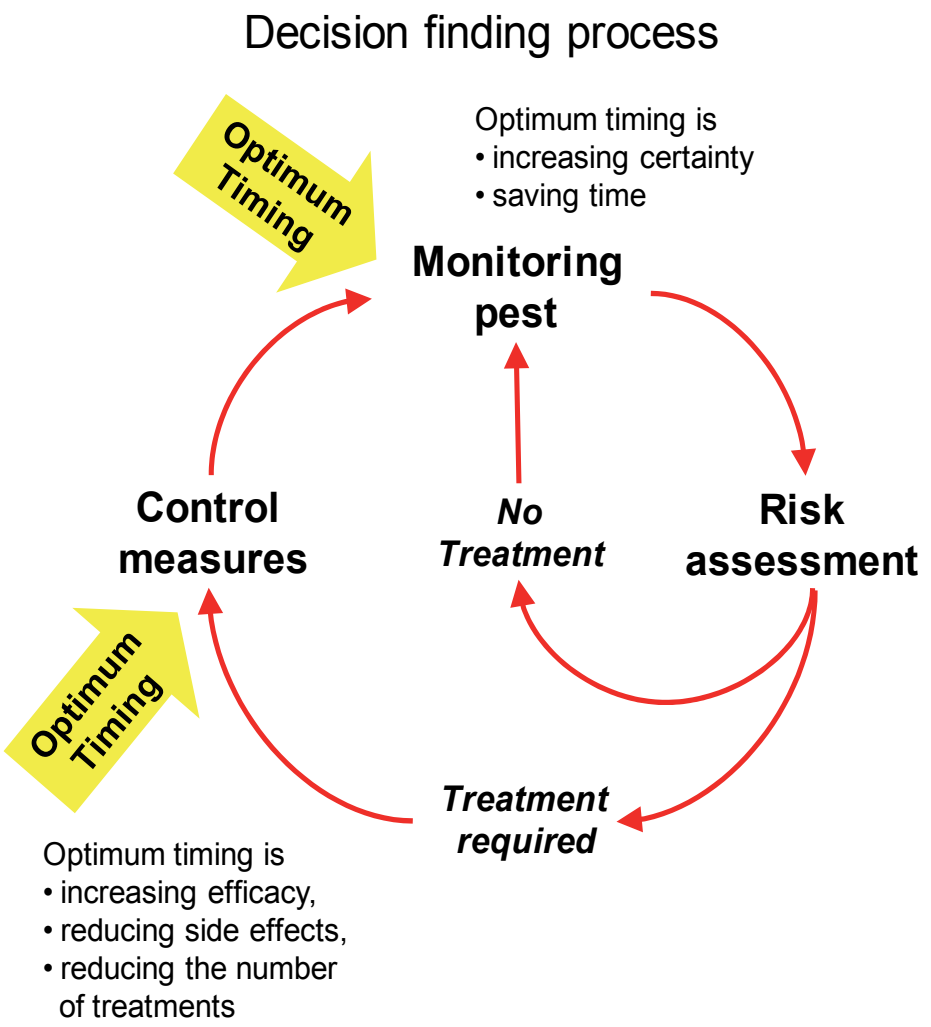


Fig. 1. Importance of precise timing during decision finding process in plant protection

Applying time-varying distributed delay approaches, phenology-models were developed driven by solar radiation, air temperature and soil temperature on hourly basis. Relationships between temperature and stage specific development rates for the relevant stages of the life cycles were established under controlled laboratory conditions for the most important orchard pests (Fig. 2).

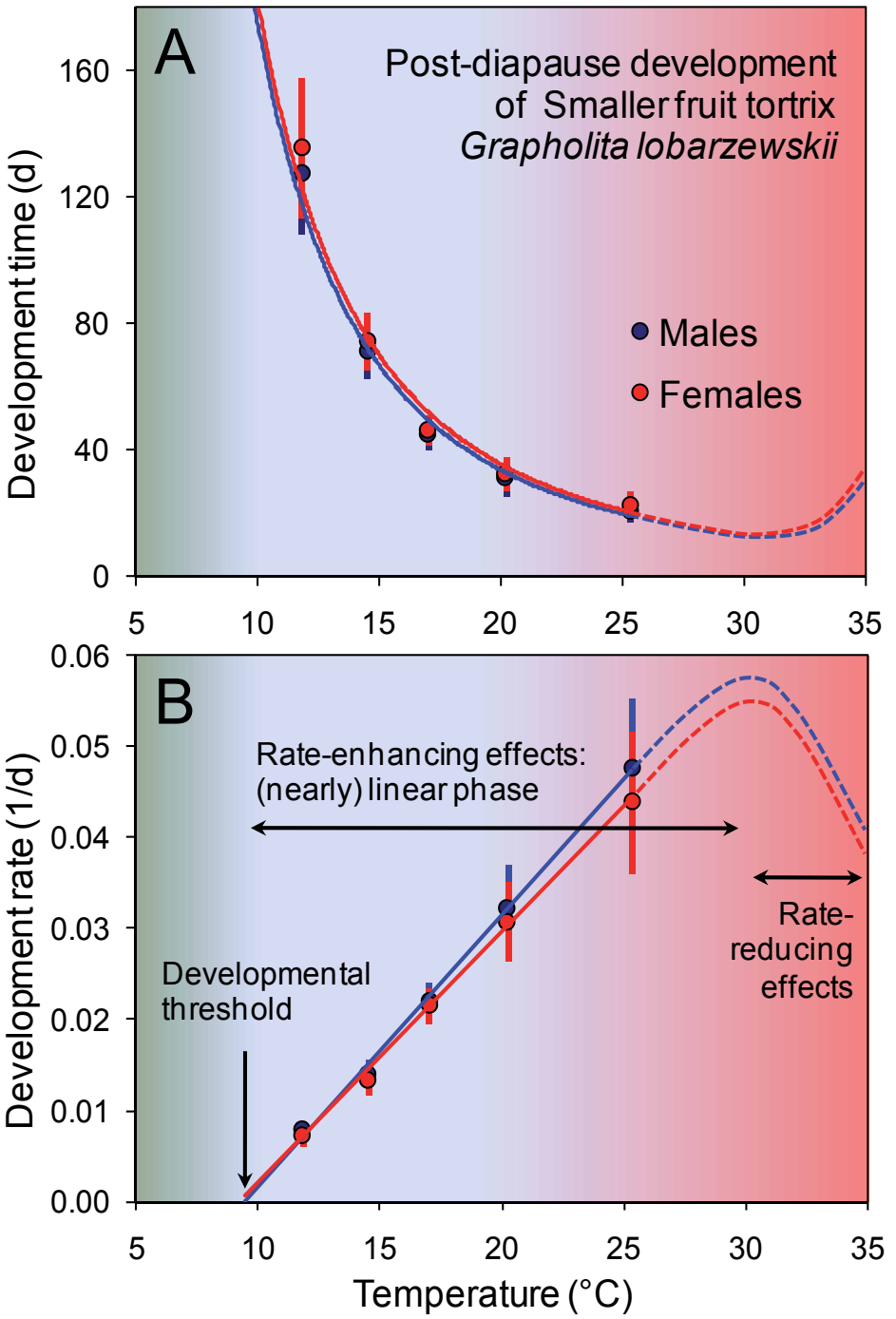


Fig. 2. Effect of temperature on post-diapause development time (A) and development rate (B) of the Smaller fruit tortrix, *Grapholita lobarzewskii* (Data after Graf et al., 1999).

Insect body temperatures in the models of SOPRA are based on studies of habitat selection of relevant developmental stages and according simulations using the driving variables and structural orchard features.
On base of local weather data, age structure of the pest populations is simulated and, based on the simulated phenology, crucial events for management activities are predicted by the SOPRA system.
Through the web-interface, the phenology is directly linked to the decision process of the fruit farmers throughout the season. Thereby SOPRA serves as decision support system for the major insect pests of fruit orchards on local and regional scale and has a wide range of possible applications in the alpine valleys and north of the Alps.

2. Single-species phenology models in SOPRA

2.1 Time-varying distributed delays

The flow of entities with variable transit times through a given process, as applicable for insect development, can be easily simulated by time-varying distributed delay models (Manetsch, 1976; Severini et al., 1990; Gutierrez, 1996).
This approach makes use of an Erlang density function to generate the frequency distribution of the individual development times, and is parameterized with the thermal constant of the specific developmental stage and its variance. An algorithm originally written by Abkin & Wolf (1976) was adapted to compute the process of aging within the different developmental stages and to continuously keep track of the age structure of the population. The changes in the age structure of the pest populations are continuously recorded by a balance of input and output from the state variables i.e. the developmental stages implemented for the single species.
In a poikilothermic development process, the mean transit time in calendar time units and its variance vary dramatically depending on temperature as exemplified here for the Smaller fruit tortrix (Fig. 2 A). Within the rate-enhancing phase of temperature, high temperatures lead to faster development, i.e. higher development rates. Low temperatures slow down biological processes until development nearly stops at the so-called developmental thermal threshold (Fig. 2 B). The relationship of process rate and temperature rises until the so-called optimum temperature is reached and decreases above this optimum due to rate-reducing or destructive effects (cf. Fig. 2), mostly at first as reversible structural damage of the enzyme systems (Somero, 1995; Willmer et al., 2000).
In order to account for these effects of temperature on biological processes, developmental time and variance are not considered as constants in the present modelling approach but as variables that are updated for each simulation step (i.e. for every hour of the season) on the basis of relationships between temperature and process rates (see below). These relationships are kept as simple as possible, which means that linear rate functions are applied wherever they give appropriate approximations (cf. Fig. 2).
Thereby, for a single delay, the ratio between the square of the mean transit time in physiological time units (i.e., the simplifying so-called thermal constant in day-degrees) and its variance specifies the order (k) of the delay and hence the number of first order differential equations required to generate the observed variability. Each of these k first order differential equations represents an age class within the simulated stage and describes the daily changes in this age class as

$$\frac{dQ_i(t)}{dt} = r(t) \times [Q_{i-1}(t) - Q_i(t)] \quad (1)$$

where $Q_i(t)$ stands for the proportion of individuals in age class i at time t. The parameter r(t) is the transition rate from age class i to age class i + 1 and corresponds to the developmental rate that is quantified with the according function describing development as function of temperature. Individuals leaving the last age class are entering the next implemented developmental stage (Fig. 3).
The algorithms implementing the time-varying distributed delay models were originally written in Pascal and later implemented with the program Delphi 6.0 (Borland, Cupertino, CA and Atlanta, GA, USA) in a MS Windows application (cf. below).

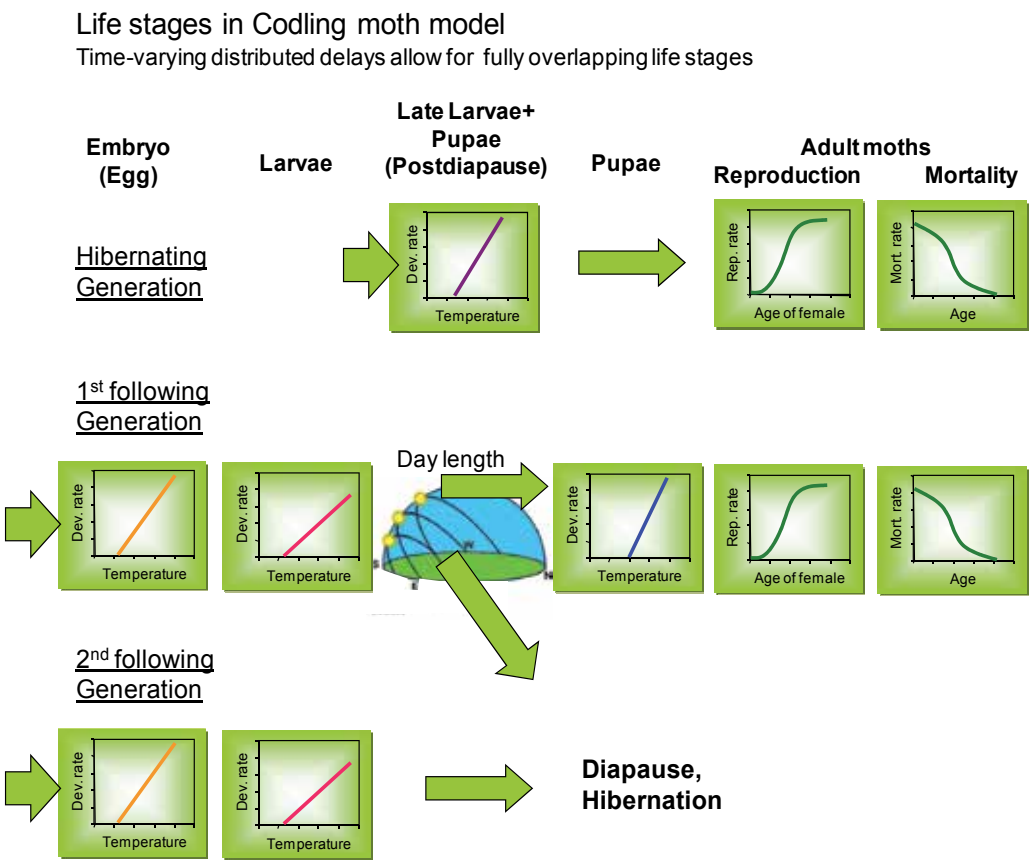


Fig. 3. Life stages implemented in the Codling moth model and schematic relationship between temperature and process rates (development rate, reproduction rate, mortality rate) for the respective stages of the life cycle. Age is always transferred to physiological time above the developmental thermal threshold to account for the temperature effect.

Species	Implemented stages of the life cycle					Temperature parameter
Rosy apple aphid (*Dysaphis plantaginea*)	Winter eggs	Juveniles	Adults	Juveniles (1. Gen.)		Air
Apple sawfly (*Hoplocampa testudinea*)	Hib. pupae*	Adults	Eggs	Larvae		Soil* Air
Smaller fruit tortrix (*Grapholita lobarzewskii*)	Hib. larvae/ pupae (M/F)*	Adults (M/F)	Eggs	Larvae	Hib. larvae (diapause)*	Inner stem* Air
Codling moth (*Cydia pomonella*)	Hib. larvae/ pupae (M/F)*	Adults (M/F)	Eggs (1. Gen.)	Larvae (1. Gen.)		Stem surface* Air
	Pupae (1. Gen.)	Adults (1. Gen.)	Eggs (2. Gen.)	Larvae (2. Gen.)	Hib. larvae (diapause)*	Stem surface* Air
Pear psylla (*Cacopsylla pyri*)	Hib. adults (M/F)	Eggs (1. Gen.)	Larvae (1. Gen.)			Air
	Adults (1. Gen.)	Eggs (2. Gen.)	Larvae (2. Gen.)			Air
Cherry fruit fly (*Rhagoletis cerasi*)	Hib. pupae (M/F)*	Adults (M/F)	Eggs	Larvae	Hib. pupae (diapause)*	Soil* Air
Apple blossom weevil (*Anthonomus pomorum*)	Hib. adults (M/F)*	Active adults (M/F)	Immigrated adults (M/F)	Eggs	Larvae	Soil/Air* Air
Summer tortrix (*Adoxophyes orana*)	Hib. larvae (M/F)*	Active larvae (M/F)	Pupae (M/F)	Adults (M/F)	Eggs (1. Gen.)	Stem surface* Air
		Larvae (1. Gen.)	Pupae (1. Gen.)	Adults (1. Gen.)	Eggs (2. Gen.)	Air
		Larvae (2. Gen.)	Hib. larvae (diapause)*			Stem surface* Air

Table 1. Species implemented in the forecasting system SOPRA with modelled stages of the life cycle and temperature driving the models. First table line of each species starts with the hibernating stage (Hib.) and following lines represent subsequent generations with the same stages below each other. F - females, M - males.

2.2 Temperature-dependent development

In SOPRA, the relationships between temperature and process rates in the single-species models are implemented with linear or non-linear functions for each stage of the life cycle depending on the nature of the best approximation (Fig. 3). All of those relationships for the relevant stages of the life cycles were established in thorough individual-based laboratory experiments under controlled conditions with a minimum of four temperature treatments for each species. Developmental rates of the stages are mostly implemented with linear functions (cf. Fig. 3). Non-linear functions are used for reproductive rates and survival of adults (e.g. Graf et al., 1996; Graf et al., 1999; Graf et al., 2001a; Schaub et al., 2005; Graf et al., 2006).

At the present, phenology models in SOPRA are established for Rosy apple aphid (*Dysaphis plantaginea*) Apple sawfly (*Hoplocampa testudinea*), Codling moth (*Cydia pomonella*), Smaller fruit tortrix (*Grapholita lobarzewskii*), Apple blossom weevil (*Anthonomus pomorum*), Summer fruit tortrix (*Adoxophyes orana*), European pear psylla (*Cacopsylla pyri*) and European cherry fruit fly (*Rhagoletis cerasi*). Recent validation of two new models extends the coverage of major pome and stone fruit pests by European red spider mite (*Panonychus ulmi*) and Plum tortrix (*Grapholitha funebrana*).

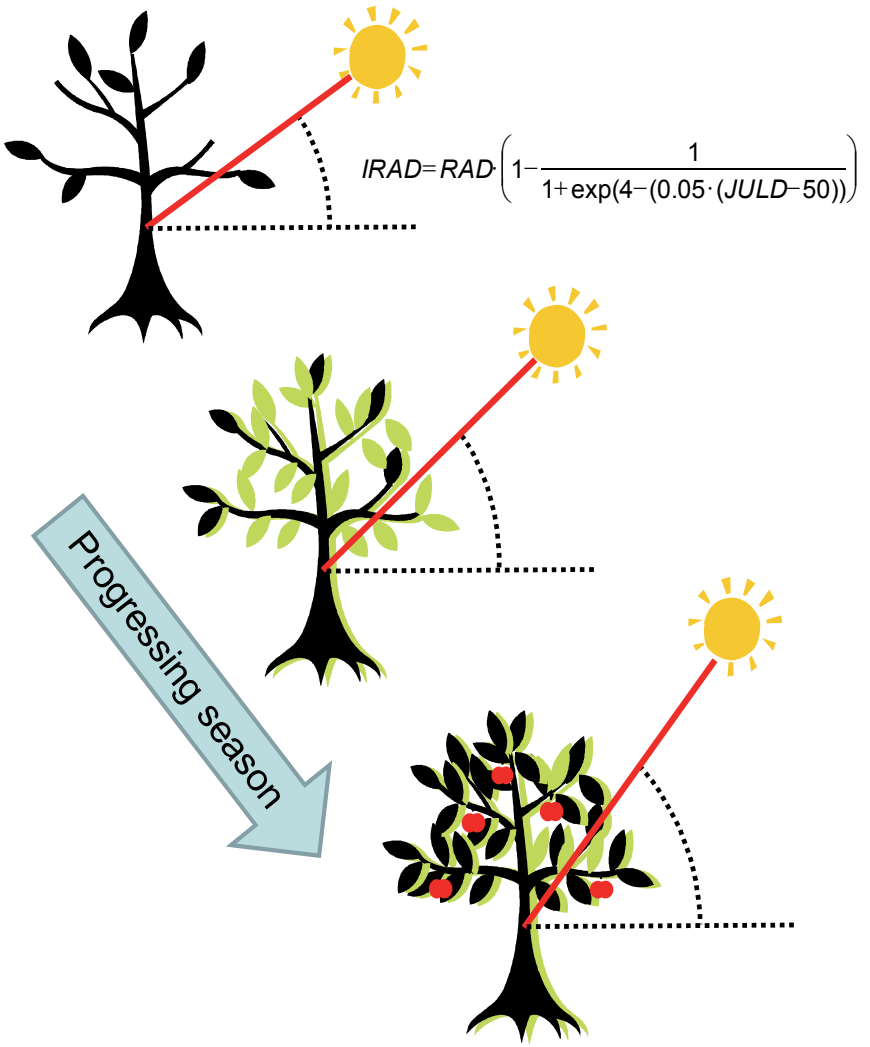


Fig. 4. Simulation of effective solar radiation (IRAD) at the plant stems with progressing season by measured solar radiation (RAD) and day of the year, i.e. Julian day (JULD).

2.3 Habitat selection and simulation of habitat temperatures

The implementation of temperatures in the models is based on intensive studies of habitat selection of the developmental stages and biophysical modelling using the three driving

variables (solar radiation, air temperature, soil temperature) and structural orchard features. Body temperatures of all implemented developmental stages are approximated by modelling habitat temperatures as close as possible (Table 1).
Soil temperature is used for post diapause development of Apple sawfly and Cherry fruit fly pupae. Stem surface temperature is implemented for hibernating larvae and pupae of the Codling moth and the Summer tortrix. Stem surface temperature is simulated from air temperature and solar radiation on base of seasonal azimuth angle of the sun and light extinction of the vegetation (Fig. 4; cf. Graf et al., 2001b). Inner stem temperature is also simulated from air temperature and solar radiation and implemented for hibernating larvae and pupae of the Smaller fruit tortrix. The remaining habitat temperatures are approximated by air temperature (Table 1).

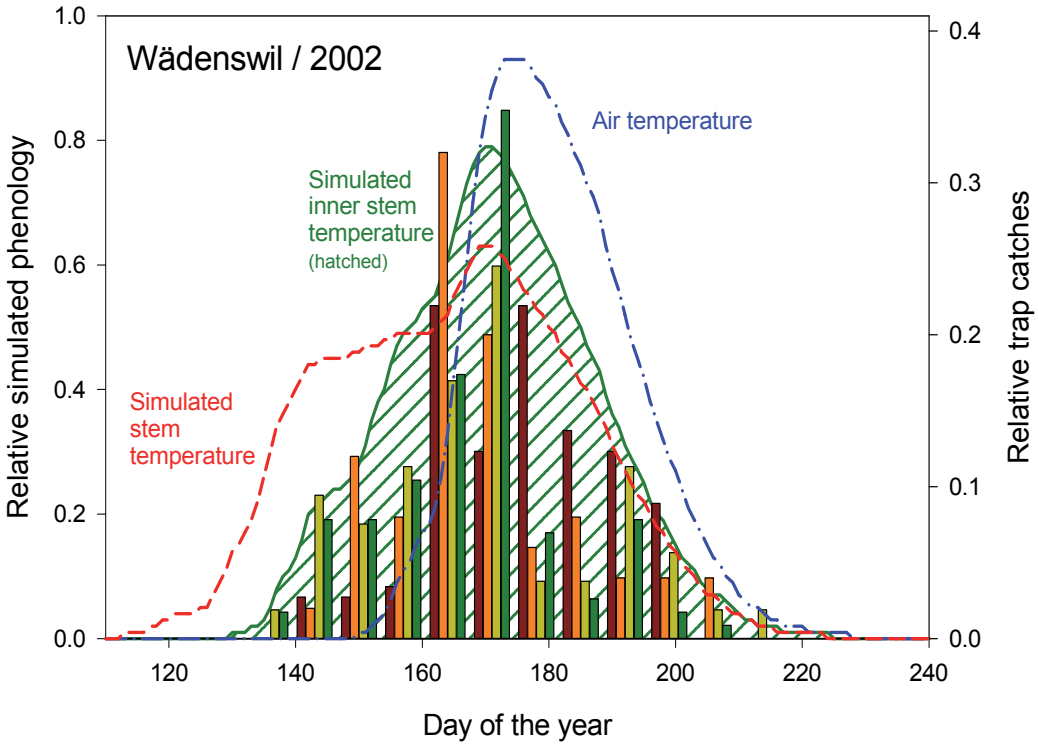


Fig. 5. Exemplified model validation in the Smaller fruit tortrix by simulated male phenology (left axis) with pheromone flight trap data (right axis).

Smaller fruit tortrix exemplifies the high importance of applying the temperature approximated to the habitat as good as possible. When comparing relative trap catches in pheromone traps, simulations of relative phenology with habitat specific inner stem temperature as applied in SOPRA (hatched in Fig. 5) shows a perfect match whereas stem surface and air temperatures lead to high deviations of phenology forecasts (Fig. 5).
For validation, simulated emergence processes from hibernation sites were first compared with emergence data from semi-field experiments. In a second step, implemented model predictions were validated with independent field observations from several years for adult

activity (e.g. Graf et al., 1996; Graf et al., 1999; Graf et al., 2001a; Schaub et al., 2005; Graf et al., 2006).

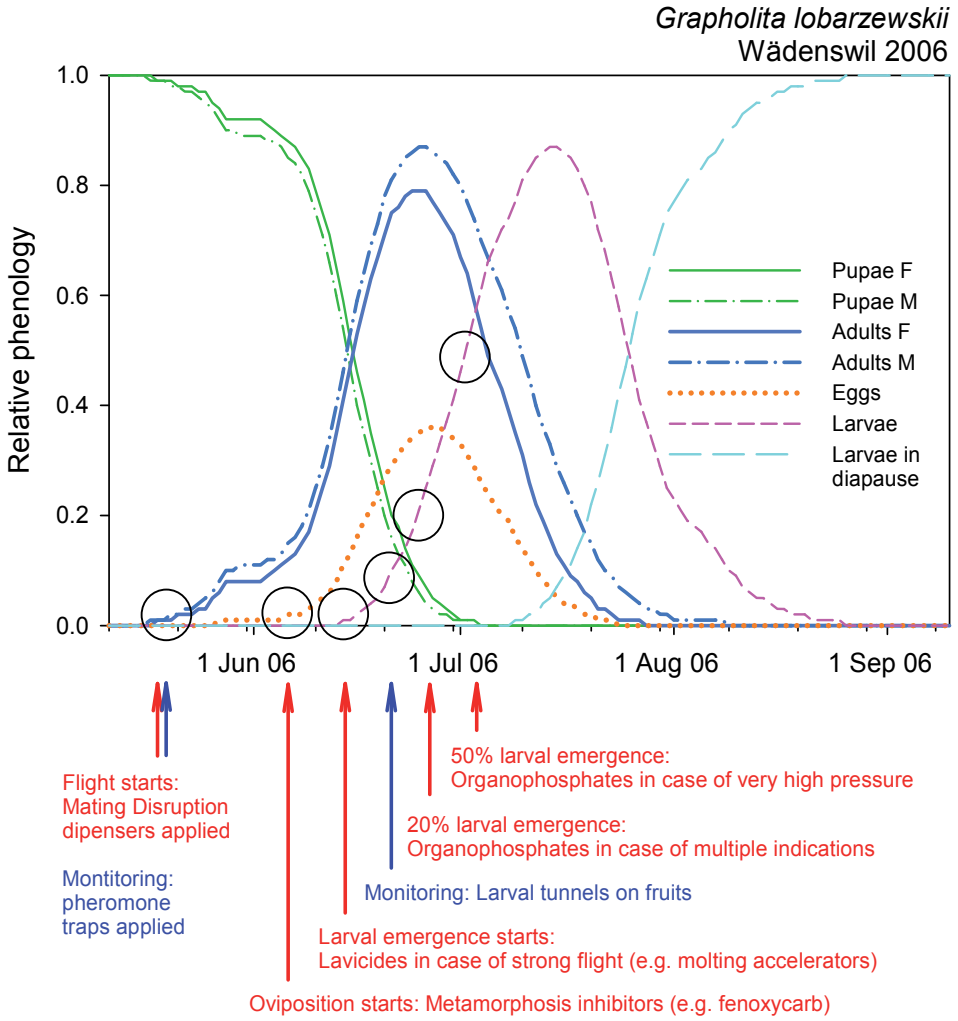


Fig. 6. Simulated relative phenology by example of the Smaller fruit tortrix with important events in the lifecycle and according suggestions. F - females, M - males.

3. Local simulation tool

The locally based user interface is designed as a common MS Windows application in order to facilitate the simultaneous use of the different species models, and to standardize the weather data input for all models. The simulation of habitat temperature is integrated in the application with a flexible weather module. Both numerical and graphical outputs are implemented. For further pests a routine is implemented to compute temperature sums for any user defined temperature threshold.

The models for each pest species or the temperature sum routine are accessed with tab controls. Weather data input text files are selected in an "open file" dialog. Check boxes allow to choose the output in numerical and/or graphical form. On base of the chosen local weather data, relative age structure of fruit pest populations is simulated and crucial events for management activities are predicted on that basis (Fig. 6). For the latter the model output is automatically interpreted in a summary table which delivers short recommendations. The decision support system on the internet platform gives even more detailed information and – besides timing for optimum monitoring or treatment – also pre- or post warning times, respectively (cf. below).

In the temperature sum routine, the user chooses air-, soil and/or stem temperature and specifies up to three different thermal thresholds and the starting day for calculation within the year. The latter also allows to calculate temperature sums from any biofix, e.g. beginning of adult flight in the field. From the numerical outputs of single species models the relative phenologies are saved and transferred to a data base for online presentation and decision support as explained below.

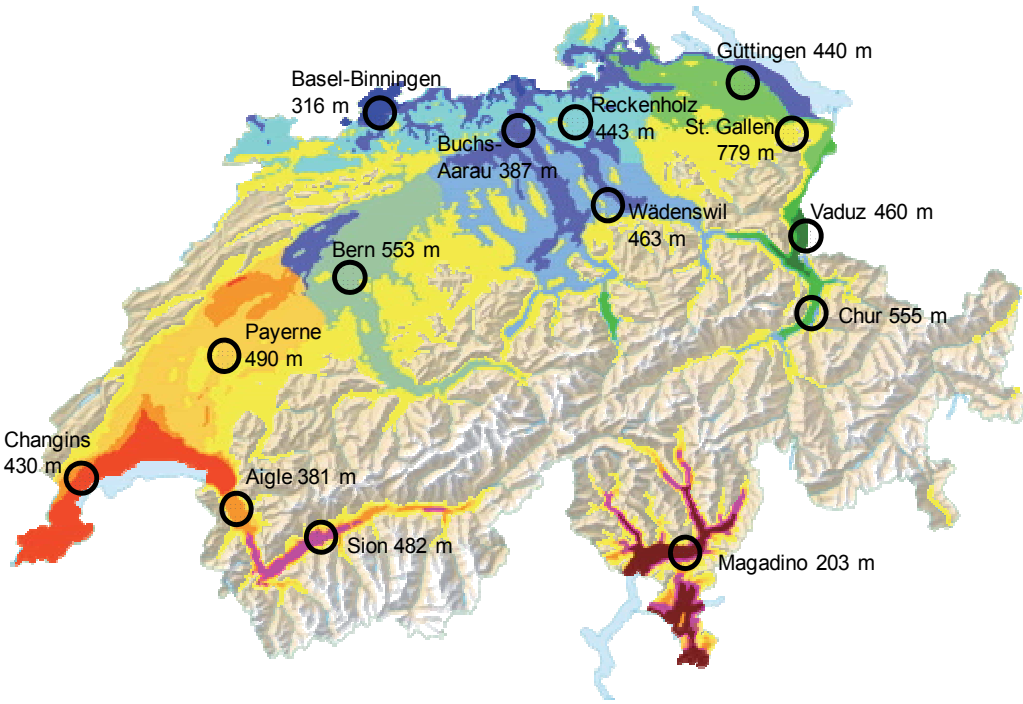


Fig. 7. Weather stations, altitude, and climatic regions that the stations are representing.

4. Weather data for simulating pest phenology

Local weather data on hourly basis (solar radiation flux density, air temperature at 2 m, soil temperature at -5 cm) are retrieved from official standard meteorological stations (MeteoSwiss) and automatically stored daily in the morning in the weather database. Ten-years means on hourly basis serve for projection and are merged with the current weather

data until the present day for simulation of phenologies throughout the entire season. Currently 14 weather stations are used to cover all climatic regions of Switzerland that are important for fruit growing (Fig. 7). The stations range from the very early Ticino valley south of the main Alpine range (station Magadino, 203 m a.s.l.), the rather early regions within the Alpine valleys (e.g. station Sion, 482 m a.s.l., station Vaduz, 460 m a.s.l.) to the late fruit growing regions above 600 m altitude in north-eastern Switzerland (station St. Gallen, 779 m a.s.l.).

5. Decision support through web-interface

Through a web-interface, the simulation results are made available to consultants and growers together with extensive information about the pest species and dynamic decision support according to the phenology in German, French and Italian, the three major official languages of Switzerland (www.sopra.info). The website is entered through the phenology forecasting part to facilitate the shortest possible way from entering the site to an overview over all pests of a certain location. Accordingly the entrance site provides a clickable map of Switzerland with the climatic regions of the 14 representative weather stations (cf. above) drawn in colour shades with relief, rivers, and lakes included.

By clicking a certain point on the map the user is led to a tabled overview of all pests at that location which is centred on the current period (Fig. 8 A). The table with the present alert status can be dynamically scrolled through the entire year or zoomed out for overview. Table cells with the species/day combination provide a colour code for monitoring (blue) and control measures (red) that is unified throughout the site. Additionally the code is divided into pre- and/or post warning phases of the announced events (light blue and red) and optimum times for certain monitoring and control measures (dark blue and red). For local differences in phenology, e.g. at southern exposed locations, reference links are provided to the earlier and later neighbouring regions (Fig. 8 A).

Clicking on the table cells leads to the core of the decision support system with graphical output of the relative age structure of the pest populations and according verbal interpretation. A chart shows the proportions of life stages along the time line (Fig. 8 B). Fig. 6 gives an example for such a visualisation of relative age structures for the Smaller fruit tortrix. The chart is scrollable throughout the year and has three zoom steps and controls for shifting the current day for which the interpretation is given as decision support. The latter is divided into monitoring and control measures indicated by the colour code mentioned above. The interpretation is referred directly to the age structure of the pest and accordingly announces crucial events for certain management activities (Fig. 8 B). Preference is given to environmentally friendly and sustainable measures like pheromone mating disruption or insect growth regulators although all other options of control measures are explained as well. The recommendations give reference to a separate part of the web site with richly illustrated information to the pest species biology and development, to monitoring methods and economic thresholds, anti-resistance strategies as well as to the list of suggested plant protection measures along with additional information on modes of action, doses, toxicity, restrictions etc.

Additionally to the entrance by the map of climatic regions, below we provide a tabulated overview of the current-day alert status for all climatic regions and species (Fig. 9). This is especially useful for quick notion of important stages in the life cycle and according events, e.g. for daily visit by consultants. Table cells with the region/species combination lead directly to the graphical output and according verbal interpretation.

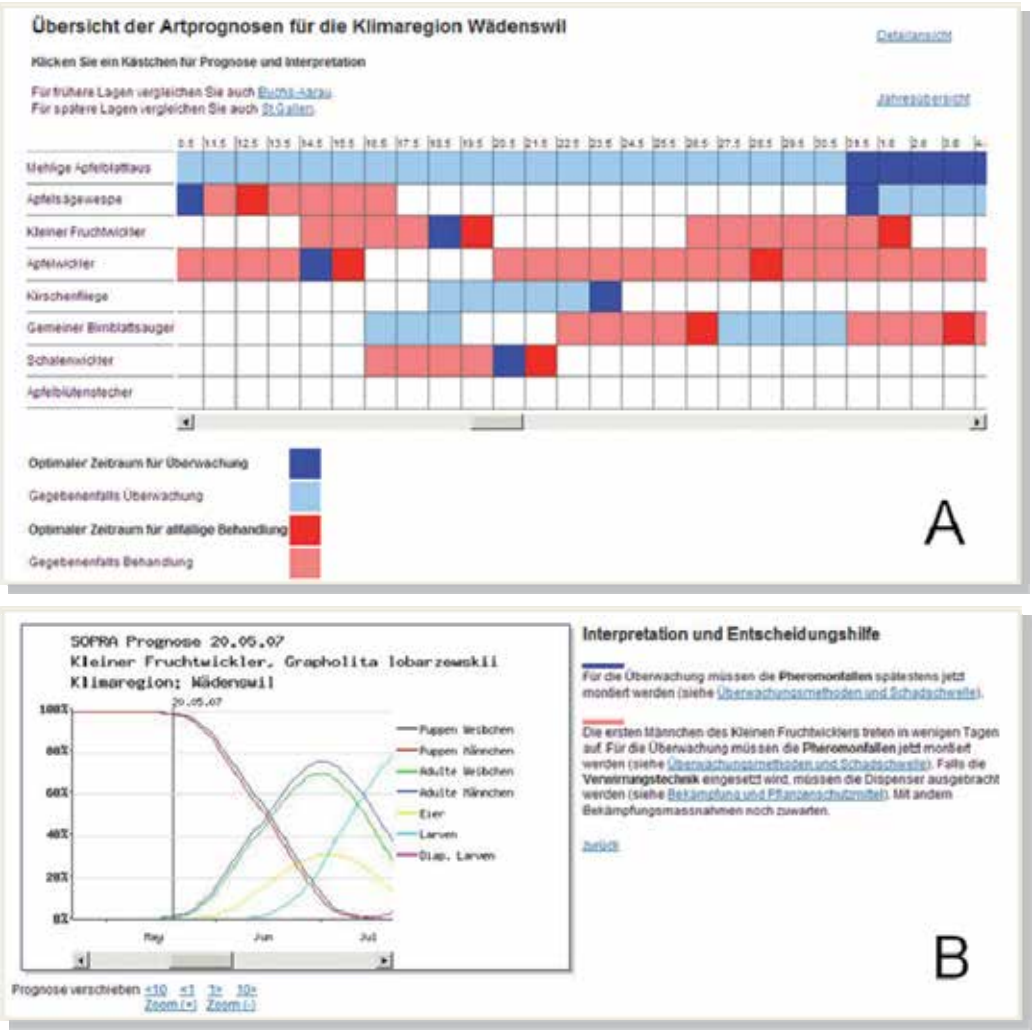


Fig. 8. Screen-shots of the web-based decision support system SOPRA with a tabled overview of all pests at one location (A) and graphical output of the relative age structure of the pest populations (B, left) and according verbal interpretation (B, right) of monitoring options (blue) and plant-protection measures (red).

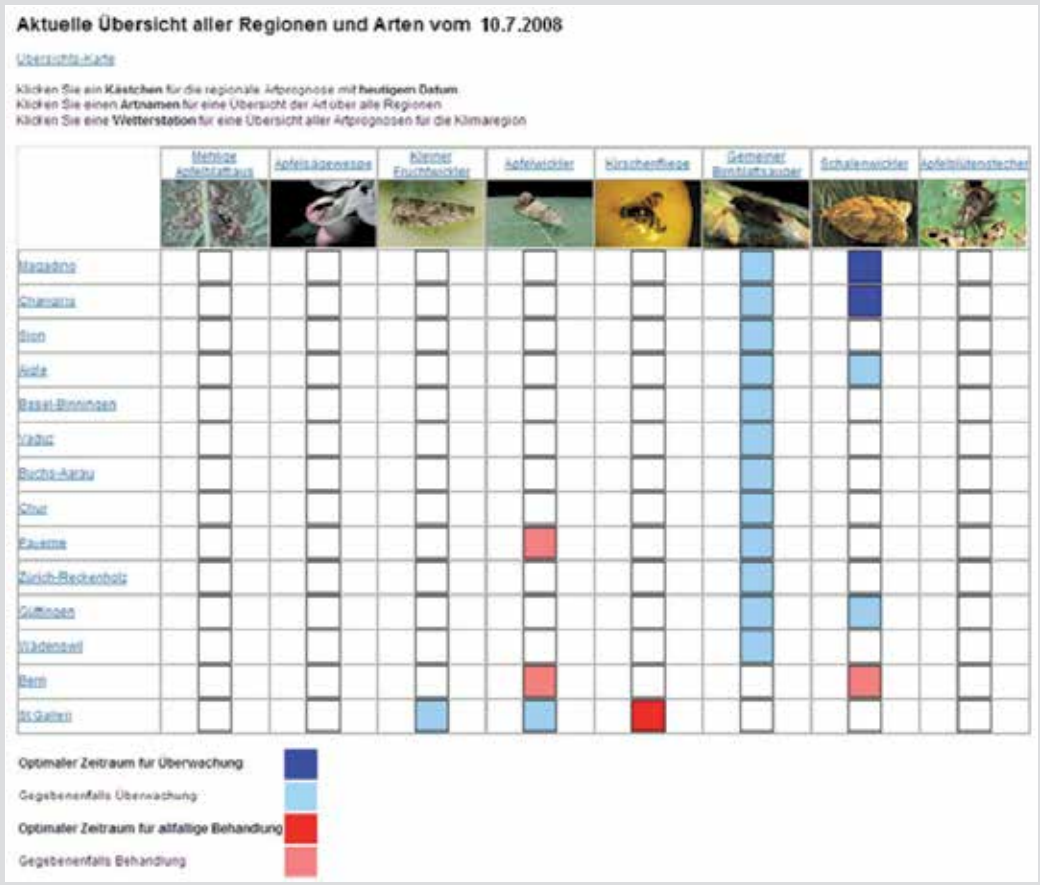


Fig. 9. Current-day alert status for all climatic regions and species in the web-based decision support system SOPRA with colour code for of monitoring options (blue) and plant-protection measures (red). Dark colours stand for optimum times for intervention; light colours pre- and post warning times.

Furthermore for each species an overview table of all climatic regions is available that is sorted by phenology. Those tables are designed especially for consultants that need a more overall impression of the phenology and its variability between certain regions. Again the table with colour coded alert status can be scrolled through the entire year or zoomed out for overview. Table cells with the region/day combination also lead directly to the core of the decision support system with graphical output and according verbal interpretation.

6. Conclusions and outlook

Major obstacles for efficient use of simulation models by extension advisors and consultants are the diversity of model approaches, the missing standards for data input and output, and the lack of a user-friendly interface (Rossing et al., 1999; van der Werf et al., 1999). We solved the main problems with integration of all target species in one flexible and extendable simulation tool that was written as a common MS Windows application. As a

further advantage the local simulation tool of SOPRA can easily be expanded for additional pests since it has an open structure and requires a limited number of simple parameters which can be established by means of standard experiments or – supported by a thorough validation – even from the literature.

The web application allows to be used both by consultants and growers which was one of the most important aims of our project. Growers reach the information about the pest situation in their area with only one click and the current decision support with just a second click. On the other hand, consultants are provided with overview tables that allow conclusions on a countrywide scale. To keep the system as simple and concise as possible, we did not include a site specific registration of orchards which can be an advantage in field crops but not necessarily in tree fruit growing.

Spatial resolution of forecasts of course depend on the availability of locally recorded temperature and radiation data. In Switzerland the governmental extension services maintain a growing network of small weather stations for scab, downy mildew and fire blight warnings. Nevertheless, we restrict our data to the official meteorological stations due to their much better accuracy – especially of air temperature measurements (cf. Sacchelli et al., 2008). Phytopathological forecasts also depend on precipitation data that are more influenced by relief and other local characteristics than the temperature data applied in our system. Although a finer network of stations could lead to a more distinct differentiation of locations, at the present stage the 14 representative climatic regions used in SOPRA seem to provide sufficient information on the Swiss scale. Nevertheless, interpolation of precise weather data could improve the local application in future.

SOPRA has been successfully applied now for about nine years as a reliable tool for recommendations in apple pests on local and regional scale in Switzerland and also in southern Germany. Since 2007, the system was also implemented for the major pests in cherry and pear, European cherry fruit fly and European pear psylla. In 2008, Apple blossom weevil and Summer fruit tortrix were added. The recently validated models for the European red spider mite and Plum tortrix are intended to be online during 2011 and further extensions are planned for the future.

By proper timing of monitoring and pest control measures the decision support by SOPRA increases the efficacy of pest management and reduces side effects. It advances environmentally friendly and save control measures like mating disruption or insect growth regulators since those measure especially depend on precise timing. SOPRA provides an important contribution to integrated fruit production being a decision making process whereby growers select from a variety of tactics to keep pests below economic damage thresholds, while minimizing environmental impact.

7. Acknowledgments

Philippe Blaise (ETH Zurich) is acknowledged for his contribution to the first versions of the simulation platform. Roger Hess and Paul Wälti (MSI AG, Buchs) contributed with coding for the web application. Mauro Genini thankfully corrected the French and Italian versions of the web application. Anja Lahusen, Thekla Ackermann, Lucie Franck, Sandra Noser and Sonja Stutz helped in the laboratory and with field validation experiments. Claudia Daniel and Eric Wyss (FiBL Frick) provided breeding material and were involved with validation of the cherry fly model.

8. References

Abkin, M. H. & Wolf, C. (1976). Distributed delay routines. Dept. Agric. Econom., Michigan State Univ., East Lansing, Class Document 8, 1-13.

Blommers, L. H. M. (1994). Integrated pest management in European apple orchards. *Annual Reviews of Entomology*, 39, 213-241.

Dent, D. R. (1995). Integrated Pest Management. Chapman & Hall. London.

Gutierrez, A. P. (1996) Applied Population Ecology: A Supply-Demand Approach. John Wiley & Sons, New York, USA.

Graf, B.; Höhn, H. & Höpli, H. U. (1996). The apple sawfly, *Hoplocampa testudinea*: a temperature driven model for spring emergence of adults. *Entomologia Experimentalis et Applicata*, 78, 301-307.

Graf, B.; Höhn, H. & Höpli, H. U. (1999). The smaller fruit tortrix, *Grapholita lobarzewskii*: predicting the phenology of adult emergence. *Entomologia Experimentalis et Applicata*, 93, 299-304.

Graf, B.; Höhn, H. & Höpli, H. U. (2001 a). The apple sawfly, *Hoplocampa testudinea*: Temperature effects on adult life-span and reproduction. *Entomologia Experimentalis et Applicata*, 98, 377-380.

Graf, B.; Höhn, H. & Höpli, H. U. (2001 b). Improving the prediction of adult codling moth (*Cydia pomonella* L.) emergence in a natural environment. *IOBC/wprs Bulletin*, 24, 127-132.

Graf, B.; Höpli, H.U.; Höhn, H. & Samietz, J. (2006). Temperature effects on egg development of the Rosy Apple Aphid and forecasting of egg hatch. *Entomologia Experimentalis et Applicata*, 119, 207-211.

Kogan, M. (1998). Integrated pest management: Historical Perspectives and Contemporary Developments. *Annual Reviews of Entomology*, 43, 243-270.

Manetsch, T. J. (1976). Time-varying distributed delays and their use in aggregative models of large systems. *IEEE Transactions on Systems Man and Cybernetics*, 6, 547-553.

Morgan, D. & Solomon, M. G. (1996). Pest-Man: a forecasting system for orchard pests. *Acta Horticulturae*, 416, 273-277.

Norton, G. & Mumford, J.D. 1993. Decision Tools for Pest Management. CAB International, Wallingford, U.K.

Pedigo, L. P. & Rice, M. E. (2005). Entomology and Pest Management. Prentice-Hall, Upper Saddle River, New Jersey, USA.

Rossing, W. A. H.; van der Werf, W. & Leeuwis, C. (1999). Quality of modelling in fruit research and orchard management: an introduction to the workshop. *Acta Horticulturae* 499: 147-150.

Sacchelli, M.; Siegfried, W. & Samietz, J. (2008). Temperature accuracy of agricultural weather stations and consequences for phenology modeling. *Acta Horticulturae*, 803, 105-110.

Severini, M.; Baumgärtner, J. & Ricci, M. (1990). Theory and practice of parameter estimation of distributed delay models for insect and plant phenologies. In: *Meteorology and Environmental Sciences*, Guzzi, R.; Navarra, A. & Shukla, J. (Eds.), 676–719 pp, World Scientific Publishing, Singapore.

Schaub, L.; Graf, B. & Butturini A. (2005). Phenological model of pear psylla *Cacopsylla pyri*. *Entomologia Experimentalis et Applicata*, 117, 105-111.

Somero, G. N. (1995). Proteins and Temperature. *Annual Reviews of Physiology*, 57, 43-68.

Van der Werf, W.; Leeuwis, C. & Rossing, W. A. H. (1999). Quality of modelling in fruit research and orchard management: issues for discussion. *Acta Horticulturae*, 499, 151-159.

Welch, S. M.; Croft, B. A.; Brunner, J. F. & Michels, M. F. (1978). PETE: An extension phenology modeling system for management of multi-species pest complex. *Environmental Entomology*, 7, 482-494.

Willmer, P.; Stone, G. & Johnston, I. (2000). Environmental Physiology of Animals. Blackwell-Science, Oxford, U.K.

19

The Volatility and Growth Trend of Field Production Mechanization Level in China - A Positive Analysis Based on Hodrick-Prescott Method

Huang Y. X., Yang X. H., Huang S., Zhu R. and Zhang Z.
Northwest A&F University
China

1. Introduction

Agricultural machinery is the important material base for developing modern agriculture, and agricultural mechanization is a symbol of agricultural modernization. It is an urgent demand for improving agricultural productivity, increasing farmers' income, ensuring agricultural production ability and promoting rural development to promote agricultural mechanization level increasing sustainably, stablely and high-speedily(Yang, 2005; Zhang and Gao, 2009; Ju and Wang, 2009). During the 1940s to 1960s, Most of the developed countries had realized the mechanization of field production and, around the 1960s, had realized all-sided mechanization successively. Usually, mechanization of field production is used instead of agricultural mechanization (similarly hereinafter). During the developing history in the past 60 years, the Chinese agricultural mechanization has experienced the following process: start-up stage, preliminary development stage and fall, recovery and steady development stage. In 2008, the level of mechanical planting, mechanical sowing and mechanical harvesting are respectively 62.9%, 37.4% and 31.2%. This shows that Chinese agricultural production mode is changing constantly and the mechanization production methods are gradually playing a leading and dominant role in agricultural production mode. However, compared with the national economic and social development needs, the level of Chinese agricultural mechanization is still low and this is becoming the main sticking point to construct and develop modern agriculture. Especially, the low level of mechanical sowing and mechanical harvesting restricted the whole level of Chinese agricultural mechanization to develop rapidly. Due to the influence of social and economic development level, technical conditions, market environment, policy system and natural conditions, the development of field production mechanization has both the growth feature over time and certain volatility feature. The development of field production mechanization has experienced several downturn periods and these periods have adverse impact on improving agricultural labor productivity, guaranteeing the agricultural production ability and promoting rural development. Along with the agricultural mechanization and the sustainable development of national economic, ensuring that field production mechanization develops continuously, stably and rapidly plays an important role in

safeguarding the national food security, promoting farmers' income, improving labor productivity and promoting rural development(Yang, 2005; Zhang and Gao, 2009).

Seeking an optimal path for Chinese agricultural mechanization to develop continuously and steadily, we have to consider the volatility of field production mechanization development as well as the potential ability to grow in the future. At present, there are few researches on the volatility of field production mechanization development. The level of mechanical sowing and mechanical harvesting reflect the characteristics of field production mechanization in China from different aspects. Considering that the low level of mechanical sowing and mechanical harvesting as the main factor that leads to the low overall level of Chinese agricultural mechanization, this article uses Hodrick-Prescott (HP) technique and GM (1, 1) model to analyse the volatility and growth trend of field production mechanization in China from 1973 to 2008 aiming to get the volatility feature and development potential and provide theoretical basis for promoting agricultural mechanization and developing modern agriculture.

2. The analysis of volatility of field production mechanization level based on HP technique

2.1 Research methods and data

As shown in Fig 1, the development of mechanical sowing and mechanical harvesting can be summarized as "growth in volatility" or "reduction in volatility". They are the interact results of two factors: long-term trends and short-term volatility. At present, the main methodologies that used to analyse the volatility measurement of economic problems are velocity method, residual method and HP technique. Compared with velocity method and residual method, HP technique possesses the characteristics of perfect theory, using flexible use and better fitting effect. Considering the complexity of field production mechanization in China (the coexistence of growth and volatility), this article selected HP technique to measure the volatility.

Since Hodrick and Prescott (1981) used HP technique to analyse the economic cycle, this method has been used in other fields. The basic principle of HP technique is: assuming that time series Y_t is combined by trend components Y_t^T and volatility components Y_t^C and then the time series is:

$$Y_t = Y_t^T + Y_t^C \quad (t=1, 2, 3,\cdots, T) \tag{1}$$

where t is the sample size.

HP filter method is to estimate the least value of the following formula:

$$\sum_{t=1}^{T}(Y_t - Y_t^T)^2 + \lambda\sum_{t=1}^{T}[(Y_{t+1}^T - Y_t^T) - (Y_t^T - Y_{t-1}^T)]^2 \tag{2}$$

where the parameter λ is the penalty factor controlling the smoothness. And this parameter requires to be given in advance. The greater the parameter λ is, the smoother the estimate trend line is, whereas the bender. For annual data, the parameter λ mainly has two kinds of value, 100 and 6.25. When using 6.25 to filter, the trend line reflects the volatility more meticulously, and it can reflect the large scale change, as well as smaller annual ups and downs. Here, λ=6.25 is used in HP filtering analysis.

Short-term volatility, described by mutation rate RV and its computation formula is:

$$RV = Y_t^C / Y_t^T \tag{3}$$

where RV reflects the deviation amplitude of economic variables to long-term trend in a specific time.

In order to analyze the China field production mechanization level volatility and growth trend between 1973 and 2008, it is necessary to get the data of mechanical sowing and mechanical harvesting. Here, data from 1973 to 1999 is from website of China Agricultural Mechanization Information Network (which is managed by Agricultural Mechanization Management of China's Ministry of Agriculture). Data between 2000 and 2008 is from China Agricultural Yearbook. Mechanical sowing and mechanical harvesting calculating methods are as follows:

$$Mechanical\ Sowing\ Level(\%) = \frac{Mechanical\ Sowing\ Area}{The\ Actural\ Total\ Sown\ Area} \tag{4}$$

$$Mechanical\ Harvesting\ Level(\%) = \frac{Mechanical\ Harvesting\ Area}{The\ Actural\ Total\ Sown\ Area} \tag{5}$$

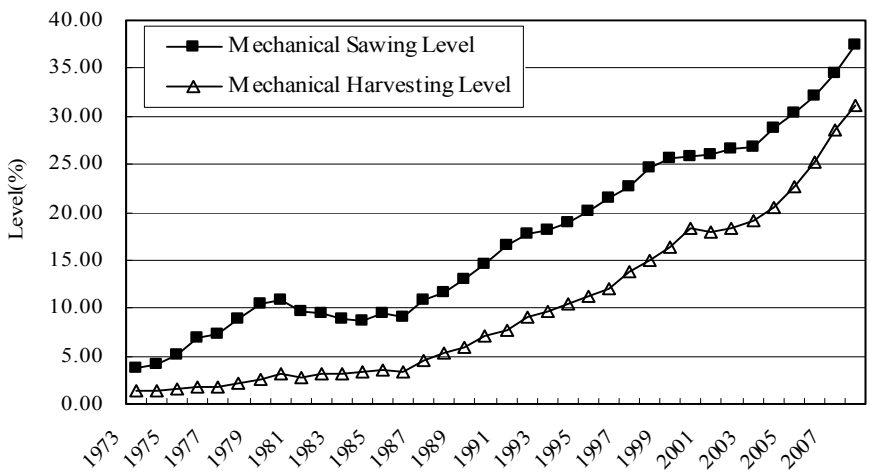


Fig. 1. The development of mechanical sowing and mechanical harvesting from 1973 to 2008 in China

2.2 The volatility analysis of mechanical sowing and mechanical harvesting

To investigate the fluctuating features of the mechanical sowing level and mechanical harvesting level, the long-term trends and short-term volatilities of the mechanical sowing level and mechanical harvesting level should be separated by HP filter. Therefore, this paper uses the software of Eviews6.0 to calculate both of the filtering results.

Table 1 shows the HP trend value of the mechanical sowing level and mechanical harvesting level in China from 1973 to 2008, which is the trend components of mechanical sowing level

and mechanical harvesting level, reflecting the endogenous, stable aspects that could be used in economic forecasting.

Fig 2 and Fig 3 show the smooth trend of the long-trend volatility of mechanical sowing level and mechanical harvesting level. From tendency sequence we know that the mechanical sowing level and mechanical harvesting level are overall on the rise from 1973 to 2008. Among them, the average annual growth of the mechanical sowing level is 0.96% while the mechanical harvesting level is 0.85%. Those figures show that the feature of volatility of the mechanical sowing level and mechanical harvesting level are obvious.

Using formula (3), we could compute the mutation rate of the mechanical sowing level and mechanical harvesting level in China from 1973 to 2008. According to periodic wave theory and the feature of "peak - valley - peak" , when the difference of wave amplitude of a cycle is bigger than 5% and the interval time is more than 3 years, we regard this cycle as an integrated cycle. It should be noted that according to different standards to define an integrated cycle the results will vary. According to this, we get the cycle of the mechanical sowing level and mechanical harvesting level in China from 1973 to 2008 and the results are shown in table 3 and table 4.

Judging from the mechanical sowing level, there were six fluctuation cycles from 1973 to 2008 and the average interval time is 6 years while the longest one is 10 years and the shortest one is 4 years. The interval time is growing from short to long. The average amplitude is 10.44% while the largest one is 17.87% and the smallest one is 5.55%. The amplitude is becoming smaller and this means that the stability of fluctuation of mechanical sowing level is becoming strengthened.

Judging from mechanical harvesting level, there were five fluctuation cycles from 1973 to 2008 and the average interval time is 7.2 years while the longest one is 8 years. Except the period from 1989 to 1992, the interval time of the other cycles is 8 years and the fluctuation frequency is stable. The average amplitude is 13.228% while the largest one is 22.90%. The amplitude is becoming smaller and this means that the stability of fluctuation of mechanical harvesting level is also becoming strengthened.

Years	Sowing level	Harvesting level	Years	Sowing level	Harvesting level	Years	Sowing level	Harvesting level
1973	3.45	1.29	1985	9.35	3.57	1997	22.82	13.76
1974	4.51	1.42	1986	9.85	3.96	1998	23.92	14.94
1975	5.61	1.56	1987	10.70	4.52	1999	24.82	16.04
1976	6.74	1.74	1988	11.80	5.23	2000	25.51	17.00
1977	7.82	1.96	1989	13.09	6.03	2001	26.10	17.80
1978	8.78	2.21	1990	14.48	6.91	2002	26.77	18.63
1979	9.49	2.47	1991	15.85	7.81	2003	27.66	19.69
1980	9.80	2.70	1992	17.10	8.72	2004	28.91	21.16
1981	9.73	2.87	1993	18.23	9.61	2005	30.49	23.09
1982	9.48	3.01	1994	19.32	10.52	2006	32.36	25.43
1983	9.22	3.15	1995	20.45	11.49	2007	34.46	28.02
1984	9.14	3.31	1996	21.63	12.58	2008	36.68	30.69

Table 1. Trend value of agricultural mechanization operation level from 1973 to 2008 in China (%)

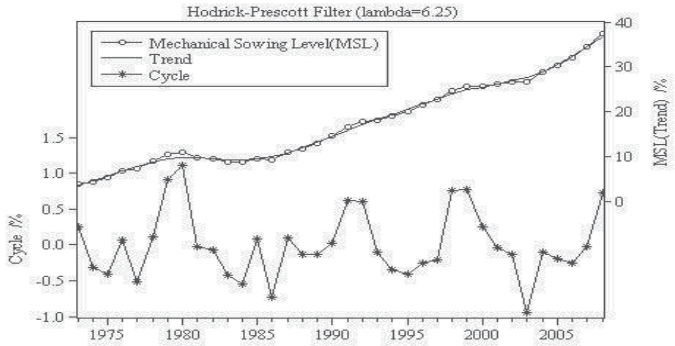


Fig. 2. Trend value of mechanical sowing from 1973 to 2008

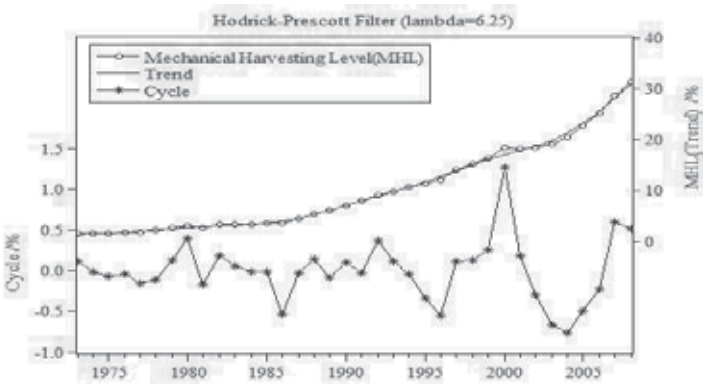


Fig. 3. Trend value of mechanical harvesting from 1973 to 2008

Years	Sowing level	Harve-sting level	Years	Sowing level	Harvesting level	Years	Sowing level	Harvesting level
1973	7.14	8.64	1985	0.90	-0.49	1997	-0.95	0.82
1974	-6.92	-1.20	1986	-7.45	-13.79	1998	3.13	0.89
1975	-7.32	-4.03	1987	0.93	-0.75	1999	3.11	1.57
1976	0.92	-2.41	1988	-1.19	2.63	2000	0.96	7.43
1977	-6.62	-8.11	1989	-0.99	-1.39	2001	-0.17	1.07
1978	1.32	-5.06	1990	0.17	1.37	2002	-0.49	-1.59
1979	9.62	5.09	1991	3.92	-0.36	2003	-3.45	-3.41
1980	11.25	14.79	1992	3.55	4.25	2004	-0.38	-3.59
1981	-0.30	-5.80	1993	-0.56	1.21	2005	-0.62	-2.13
1982	-0.79	6.29	1994	-1.82	-0.39	2006	-0.81	-0.89
1983	-4.56	1.73	1995	-2.00	-3.00	2007	-0.09	2.15
1984	-5.93	-0.42	1996	-1.16	-4.35	2008	1.97	1.66

Table 2. RV value of agricultural mechanization operation level from 1973 to 2008 in China (%)

Cycle serial number	Sowing level			Harvesting level		
	start-stop year	annual range	Amplitude(%)	start-stop year	annual range	Amplitude (%)
1	1973-1976	4	14.46	1973-1980	8	22.90
2	1977-1980	4	17.87	1981-1988	8	20.08
3	1981-1985	5	6.83	1989-1992	4	5.64
4	1986-1991	6	11.37	1993-2000	8	11.78
5	1992-1998	7	5.55	2001-2008	8	5.74
6	1999-2008	10	6.56			

Table 3. The results of Volatility features analysis of agricultural mechanization operation level from 1973 to 2008 in China

	cycle number	average annual range	Average amplitude (%)
Sowing level	6	6	10.44
Harvesting level	5	7.2	13.228

Table 4. The cycle of agricultural mechanization operation level

2.3 Consideration on promoting field mechanization level developing fast and perfectly

2.3.1 Paying attention to factors that affect field mechanization level during different periods

Through Table 3 and Table 4, the development cycle of China Mechanical Sowing Level and Mechanical Harvesting Level during the year 1973 and 2008 exists the largest difference; this is mainly because there are different development factors in different periods affecting field mechanization level. Existing researches show that the main factors that affect the field mechanization development are farm machinery management modes, rural economic development conditions, rural residents' per capita income, policy system and so on. For example, Lu etc (2008) applied the DEMATEL method to identify the factors that influence the agricultural mechanization development and finally found out six factors as key ones, namely, the farmer's income, farm machinery industry development level, farm machinery products use cost, the transfer of rural surplus labor force, the labor price, and scale of cultivated land management. Liu and Tian (2008) analyzed key factors that influenced China farm machinery equipment level. Through this analysis, the most important factor is the level of economic development level because it contributes the biggest share, and the second is the land business scale, thus expanding land management scale moderately plays an important role in promoting the level of agricultural equipment. The third is the planting structure. The factors mentioned above are those causing the volatility of agricultural mechanization, also the breakthrough of developing the agricultural mechanization.

2.3.2 Mobilizing farmers' enthusiasm of purchasing and using agricultural machinery

Mobilizing farmers' enthusiasm of purchasing and using agricultural machinery plays the most fundamental role in the development of agricultural mechanization. From 1949 to

1978, the main operation modes of agricultural machine were state-owned operation, collective ownership and collective operation. During 1970s, the operation of agricultural machine was mainly collective operation. Therefore, the improvement of the mechanical sowing level and mechanical harvesting level was the result of the inputs for agricultural mechanization from the state. To achieve the goal of 75% of agriculture, forestry, animal husbandry, sideline production and fishing mechanization till 1980, the country hand invested a lot of money. From 1971 to 1979(not including 1974) the money used to aid the agricultural mechanization was2.38 billion Yuan and the annual investment was about 0.3 billion Yuan. There was also a greater input all over the country. In addition, the price of agricultural machine had reduced 4 times from 1971 to 1978 and the reductions were respectively 15.50%, 5.4%, 10% and 13%. With the increase of the investment and the reduction of the agricultural machine price, the agricultural mechanization level had developed Greatly (Liu and Ren, 1997; Agricultural mechanization management department, 2009). The mechanical sowing level had increased from 3.7% in 1973 to 10.9% in 1980 with an annual growth rate of 1.03%; the mechanical harvesting level had increased from 1.4% in 1973 to 3.1% in 1980 with an annual growth rate of 0.24%. Overall, the mechanical sowing level and mechanical harvesting level had improved rapidly.

In the 1980s, with the system reform of rural economic, the farm machinery management mode came to a new period of the unit of family. The original farm machinery management pattern already could not adapt to the development of agricultural machinery, especially when the household contract responsibility system got popular. Farmers increased the enthusiasm to own and buy agricultural machinery when they came to know that the agricultural machinery could help them to lighten physical labor, to improve the operation quality and to enhance the work efficiency. Through this, to enhance the level of field mechanization, it is necessary to do from two aspects, namely, requirements and probabilities. Requirements emphasize whether agricultural and rural economic development require agricultural equipment, and in this point, it has reached a consensus. Probabilities emphasize whether farmers have the ability to buy agricultural equipments. When it came to the period of the unit of family, the improvement of field mechanization level mainly depended on whether farmers had ability to buy agricultural implements.

On July 1, 1994, the country cancelled the preferential policy of agricultural parity diesel. So far, the country has cancelled all the preferential policies for agricultural machine, which were implemented during the planned economy period. The development of agricultural mechanization entered the market-oriented stage. Farmers obtained the independent right of management of agricultural machinery, so they could buy agricultural machine and operate legal business independently. These policies greatly aroused the enthusiasm of the farmers for the development of agricultural mechanization and also solved the basic dynamic problems of agricultural mechanization. This promoted the development of agricultural mechanization at a high speed.

Table 5 shows the economic system and investment behavior of Chinese agricultural mechanization development in different periods. At present, it is in a mixed economy stage, and in this stage, the investment subjects include the farmers, collective and government finance. It presents the diversified structure. In the whole, the farmers' input takes the proportion of total investment in 70%, and this shows that the farmers are the most important part. Therefore, fully mobilizing farmers' enthusiasm to purchase and use agricultural machinery is the basic method to increase the level of field mechanization.

Economic System	Main Investors	Motivation	Investment Mechanism
Planned Economy	Country, Collective	State strategic goals	Administrative means, government order
Market Economy	Peasant(Peasant Household)	To meet some needs, pursue the economic benefits	Economic means, free trade
Mixed Economy	Peasant is the main body, government macro-control	Economic benefits and social benefits	Market mechanism and government functions

Table 5. Economic system and investment behavior

2.3.3 Being market-oriented, developing the agricultural machinery cooperative

Scientific and reasonable mode of agricultural machinery management is a supporting condition in developing agricultural mechanization, and also the intrinsic request to develop the modern agricultural productivity. Since the rural reform, agricultural machinery entered the market as commodity, forming resource allocation mechanism leaded by the market and the developing mechanism to pursue the economic benefit maximization, and these effectively promote the agricultural socialization and marketization. Different forms of agricultural machinery professional services firms, joint stock partnership, farm machinery professional cooperatives and farm machinery association, farm machinery users association and so on developed well in the whole country, and maintained a good momentum of development. Till now, they have become the market's main body and they are in the process of independent operation and financial self-sufficiency, self-accumulation, and self-development, fully displaying the strong vitality of farm machinery service socialization and marketization and outstanding development prospect.

Agricultural Machinery Cooperative (AMC) is the main development direction of Chinese Agricultural machinery socialization service. This is beneficial to the integrated application, large-scale promotion, moderate scale operation and industrialization (Li, 2009; Biser, 1983; Russo, 2000). 'Chinese State Council's opinion on promoting development of agricultural mechanization and farm machinery industry nicely aand rapidly' and the 'Chinese agriculture ministry's opinion on accelerating the development of agricultural machinery cooperatives have been issued Successively', which clearly require to promote the development of AMC. Table 6 shows that the number of AMC in 2007 and 2008 is respectively 4435 and 7860 and the number in 2008 is 77.23% larger than that in 2007. During this period, Agricultural machinery cooperatives in China developed at a high speed and the quantity increased substantially. This was due to the promulgation of the law on farmers' professional cooperatives and the implementation of various supporting policies. Table 7 shows statistics of the development of Chinese agricultural machinery cooperatives in 2007and 2008.

Item	2007	2008	Rate of Increase (%)
Total number of AMC	4435	7860	77.23

Data Source: The agricultural mechanization management department of Ministry of Agriculture, investigation of AMC development in 2007 and 2008.

Table 6. AMC development situation in 2008

Item	Unit	Quantity	item	Unit	Quantity
1.Total number of AMC	unit	7860	7. Machine number Where: Large and Medium Tractors Mini-Tractors Combines Rice transplanters Auxiliary farm tools	10,000 unit Individual Individual Individual Individual Individual	38.4353 54179 77902 44712 15530 217950
2. Cooperator	unit(house hold)	293327	8. General assets Where: Net value of fixed assets Facilities area	10,000 RMB 10,000 RMB khm²	1026050.09 866294.28 13284.7
3. Number of serving households	household	7537310	9. Funds supported by finance	10,000 RMB	108141.12
4. employee Where: perennial hire professional and technical personnel	Individual individual individual	267662 60695 41666	10. Quantities of AMC supported by finance	unit	3797
5. Homework service area Where: Mechanical farming Mechanical planting Mechanical harvesting Plant protecting Other	khm² khm² khm² khm² khm² khm²	11920.8 4678.4 2571.4 3364.7 836.4 1411.5	11. New AMC expected for year 2009	Individual	5466
6.Total Service revenue Where: operation services income Repair service revenue Other income	10,000 RMB 10,000 RMB 10,000 RMB 10,000 RMB	564734.59 495783.77 23657.07 42624.33	Appendix: Quantity of other farm machinery service organizations Where: Farm machinery professional associations Farm operating companies Farm machinery service (combination)	Individual Individual Individual Individual	76178 4813 1549 6447

Attach: Homework service area, Total Service revenue are the total number of 2008, others are the number of the end of year 2008.
Data Sources: The agricultural mechanization management department of Ministry of Agriculture,, investigation of farm machinery professional cooperatives development, 2007~2008.

Table 7. Statistics of China farm machinery professional cooperatives development situation in 2008(continued)

2.3.4 Improving the management efficiency of agricultural mechanization production continuously

Only when the management benefits rise, will farmers obtain economic strength and confidence to invest more money on buying agricultural machines and field production mechanization level be improved. Before 1980, China put forward slogans that we would realize the agricultural mechanization in 1980 which resulted in the bland blind development of Chinese agricultural mechanization. Such development had exceeded the development level of rural productivity during that time and could not be accepted by farmers, thus leaded to large quantities of wastes. One reason to explain this is that the government putting more focus on service other than efficiency which betrayed the economic law. Under the commodity economy conditions, profits are the major objective for farmers to purchase and use of agricultural machinery and the power for the development of agricultural mechanization (Hill, 2005; Georgeanne etc, 2009).

Many districts explored a lot to find the way to improve farm machinery management benefit. From them, improving agricultural socialization service quality and benefit continuously is an effective way to promote farmers' management benefit. Agricultural machinery cooperatives had a rapid development and service quality and benefits were improved significantly especially after the publishment of 'opinions on accelerating the development of agricultural machinery professional co-operatives'. They played even a more important role in increasing agricultural income and farmers' income. The farm machinery management service organizations must accumulate by themselves in order to increase the input of mechanization, make mechanization develop to a higher level. Therefore, it is necessary to forward the farm machinery service to the market, practising enterprise management and socialized service, improving mechanical utilization ratio and the economic efficiency, improving the level of agricultural industrialization service.

The necessary condition of increasing agricultural mechanization level is to increase the farmers' income. Under the current circumstance that the farmers' income level is low, it is possible to increase farmers' income, government subsidies and so on to solve problems of agricultural mechanization upfront input. Zhu etc (2007) studied the interdependency between agricultural mechanization development and financial input, and the results showed that the interdependency degree is related to financial investment and there is positive correlation to per capita net income of farmers. At present, China has entered the period in which the government needs to strength financial support for agricultural mechanization development and in which the dependence to finance is growing rapidly. Take '11th five-year plan' period of agricultural mechanization development as an example, the national comprehensive mechanization level came to an unprecedented increasing rate which was more than 3% from the year 2007 to year 2009. And this created a history record of rapid development. One of the most important reasons is agricultural machinery purchase subsidies increased fast year by year, and the central finance agricultural machinery purchase subsidies increased from 0.07 billion Yuan in 2004 to 15.5 billion Yuan in the year 2010, and this promoted the rapid development of operation level of agricultural mechanization.

2.3.5 Completing supporting system construction of agricultural mechanization development

Operation laws of agricultural mechanization include their own system coordination and external environment coordination. Since the publishment of People's Republic of China agricultural mechanization promotion law, through the development of 'tenth-five' plan

and '11th five-year plan' this law became perfect and improved, forming the best legal, policies and the development environment during the whole history of China. Instead of developing agricultural mechanization in isolation, China developed agricultural mechanization together with developing the agricultural machinery industry, developed advanced agricultural socialization service together with developing modern circulation technology, and developed the agricultural machinery together with establishing information construction.

Completing agricultural mechanization promotion system mainly include: agricultural machinery production, study and research, spreading and coordination system, national macro regulation and policy system, land scale management system, agricultural mechanization investment, agricultural mechanization service team construction system, financing system, agricultural machinery development of information network system, agricultural machinery service of industrialization and marketization and socialized service system, and so on. Thus, it is important to complete supporting system construction of agricultural mechanization development in order to provide system guarantee for the development of agricultural machinery (Yang etc, 2005).

3. The long-term trends and growth capacity prediction of the growth of field production mechanization level

3.1 Research method

Using HP filter method, the field production mechanization level can be separated into two parts: the value of trend and the value of fluctuation. Between them, the value of fluctuation can be used as the basis for understanding mechanization development level and the value of trend can be used as the basis for estimating and judging development of field production mechanization level in the future. At present, in the area of forecasting the field production mechanization level, the main methods we used including gray forecast method, exponential smoothing, least-square method, comperz curve method and artificial neural network technology, etc (Biondi etc, 1998; Zhang and Gao, 2009). When using the methods mentioned above, most of the data was the value of time series about field production mechanization level and it did not separate field production mechanization level into two parts of the value of fluctuation and the value of trend. Because China field production mechanization develops over time and also contains the characteristic of volatility, considering the trend components and wave components, it is difficult to forecast. In this article, the value of fluctuation was eliminated when forecasting long-term growth trend of Chinese field production mechanization level; the value of trend was used and Grey Forecasting Model was applied to predict the growth trend of Chinese field production mechanization level.

Grey System Theory was developed by Professor Deng in the 1980s (Deng, 1989). This theory focuses on studying uncertainty problems with a small number of samples or a system with poor information. Grey forecasting model refers to the GM(1,1) based on the prediction. It is not suitable for approximation complex nonlinear function, but can reflect overall development trend very well. Grey prediction model is mainly used in prediction problems with short time span, small data quantity and little fluctuation. Under the circumstances that the data quantity is small, it can make the result much more accurate. The principle and procedure of grey prediction model is shown as Follows (Deng, 1989; Wang etc, 2010):

Giving the original data sequence vectors $x^{(0)}$

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\} \tag{6}$$

Where n is the number of data in the original sequence.
Firstly, accumulate the original data sequence in order to weaken the randomness and volatility of the original data.

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\} \tag{7}$$

In the equation above,

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) \qquad (k = 1, 2, \cdots, n)$$

Step 1 Establishment of Grey Model (1, 1)

$$x^{(0)}(k) + az^{(1)}(k) = u \tag{8}$$

Here, $z^{(1)}(k)$ is generated relating to $x^{(1)}(k)$, and $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$. a is the model development coefficient. u is the gray input.
Step 2 Establishment of structural matrixes B and data vector Y_n

$$\hat{a} = [a, u]^T = (B^T B)^{-1} B^T Y_n \tag{9}$$

$$Y_n = [x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n)]^T$$

$$B = \begin{bmatrix} -0.5(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \\ -0.5(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix}$$

Here, $\hat{a}$ is a parameter that needs to be identified.
Step 3 Calculation of a and u

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -(z^{(1)}(2)) & 1 \\ -(z^{(1)}(3)) & 1 \\ \vdots & \\ -(z^{(1)}(n)) & 1 \end{bmatrix} \times \begin{bmatrix} a \\ u \end{bmatrix} \tag{10}$$

Step 4 Accumulation of the model and predication results
The analytical solution for equation (8) is:

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - u/a)^{-ak} + u/a \tag{11}$$

Here, $\hat{x}^{(1)}(k)$ is the Kth analytical solution.

Step 5 Predicted results

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{12}$$

In the equation (12), $\hat{x}^{(0)}(k)$ is the Kth predicted result by reduction.

Then, it is to get the following predicted data sequences

$$\hat{x}^{(0)} = \{\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \cdots, \hat{x}^{(0)}(n), \hat{x}^{(0)}(n+1)\} \tag{13}$$

Step 6 checking the error size, analyzing the accuracy of predicted results

GM(1,1) group includes four, five, six data and multiple data GM(1,1). In the prediction process, researchers could choose corresponding GM (1, 1) model according to the actual situation of the forecasted object (Deng, 1989).

3.2 Results and discussions

According to the formula (8) ~ (13), we established the GM(1,1) prediction model of mechanical harvesting level and mechanical sowing level based on data four, data five and data six. The main purpose is to choose the best GM (1, 1) forecasting model as the final prediction model through comparison. Among them, the GM(1,1) of data four uses the trend value of mechanical harvesting level and mechanical sowing level from 2002 to 2005, the GM (1, 1) model of data five uses the trend value of mechanical harvesting level and mechanical sowing level from 2001 to 2005 and the GM(1,1) of data six uses the trend value of mechanical harvesting level and mechanical sowing level from 2000 to 2005. The actual trend value of mechanical harvesting level and mechanical sowing level from 2006 to 2008 is used as inspection samples. Calculation results are showed in table 8 and table 9.

Table 8 shows that the average prediction error of the trend value of mechanical sowing based on data four dimension is 2.56%, obviously less than that based on data five and data six. This demonstrates that the result of GM (1, 1) model based on data four is more precise and this model can be used to forecast the trend value of mechanical sowing. The calculation equation is:

$$x(k+1) = 551.5702\exp(0.0488k) - 524.8002 \tag{14}$$

Using the equation (14), we can predict the trend value of Chinese mechanical sowing level from 2009 to 2015 and the results are shown in table 10.

Similarly, table 9 shows that GM(1,1) based on data four is the most precise one, thus we choose this model as the final model. The calculating equation is:

$$x(k+1) = 235.5554\exp(0.0800k) - 216.9254 \tag{15}$$

Using the equation (15), the trend value of mechanical harvesting level from 2009 to 2015 can be predicted, the results are also shown in table 10.

Year	Actual trend Value/%	Four dimension GM(1,1)		Five dimension GM(1,1)		Six dimension GM(1,1)	
		Predicted Value /%	Error (%)	Predicted Value /%	Error (%)	Predicted Value /%	Error (%)
2006	32.36	31.97	1.21	31.71	2.01	31.43	2.87
2007	34.46	33.57	2.58	33.13	3.86	32.69	5.14
2008	36.68	35.25	3.90	34.61	5.64	34.01	7.28

Table 8. Predicted mechanical sowing value and analysis of errors

Year	Actual trend Value /%	Four dimension GM(1,1)		Five dimension GM(1,1)		Six dimension GM(1,1)	
		Predicted Value /%	Error (%)	Predicted Value /%	Error (%)	Predicted Value /%	Error (%)
2006	25.43	24.94	1.42	24.65	3.07	24.35	4.25
2007	28.02	27.02	3.57	26.50	5.42	26.02	7.14
2008	30.69	29.27	4.63	28.49	7.17	27.80	9.42

Table 9. Predicted Mechanical Harvesting Value and Analysis of Errors

Year	Mechanical Sowing Trend Value /%	Mechanical Harvesting Trend Value /%
2009	36.97	31.71
2010	38.82	34.35
2011	40.76	37.21
2012	42.80	40.31
2013	44.94	43.66
2014	47.18	47.30
2015	49.54	51.24

Table 10. Predicted Trend Value of Mechanical Sowing and Mechanical Harvesting

4. Conclusions

Mechanical sowing level and mechanical harvesting level are important indexes that reflect development level of China's field production mechanization. It is important to research on its volatility and growth trend for promoting the development of mechanical sowing level and mechanical harvesting level in a stable and high-speed way. HP filtering method is adapted to analyse the fluctuations problems of mechanical sowing level and mechanical harvesting level and based on this, GM(1,1) is used to predict the growth trend of Mechanical sowing level and mechanical harvesting level. Results show that the mechanical sowing level and mechanical harvesting level from 1973 to 2008 China was overall rising; mechanical sowing level includes six fluctuation cycles and the average fluctuating interval is 6 years, average amplitude is 10.44%; Mechanical harvesting level contains five fluctuation cycles and the average fluctuating interval is 7.2 years, average amplitude is

13.228%; The amplitude of mechanical sowing level and mechanical harvesting level is reduced, and the growth stability of both is gradually strengthened; The factors that affect the development of field production mechanization level include farm machinery management modes, rural economic development condition, rural residents' per capita income, policy system, etc. In order to prevent mechanical sowing level and mechanical harvesting level from fluctuate greatly and promote both development at a high speed, this article put forward: attention should be focused on factors that affect production field mechanization development during different periods; enthusiasm of the farmers to purchase and use agricultural machinery should be mobilized; guided by the market, agricultural professional cooperatives should be developed quickly; management benefits of agricultural mechanization production should be improved continuously; the construction of supporting system for agricultural mechanization development should be promoted constantly. In addition,GM(1,1) is used to predict the growth trend of mechanical sowing level and mechanical harvesting level and the predicted results show that by 2015, trend value of Chinese mechanical sowing level and mechanical harvesting level will be 49.54% and 51.24% respectively.

5. References

Agricultural mechanization management department. (2009). Technology development report of agricultural mechanization in China from 1949 to 2009. China's agricultural science and technology press, pp: 3-30.

Arnon, I. (1981). Modernization of agriculture in developing countries: resources, potentials and problems. New York, John Wiley & Sons, pp: 565.

Andrea Harris; Murray Fulton. (2000). Farm machinery co-operatives in Saskatchewan and Quebec. Center for the study of Cooperatives, University of Saskatchewan.

Bagheri, N and Moazzen, S. A. (2009). Optimum strategy for agricultural mechanization development in Iran. Journal of Agricultural Technology, Vol.6, No.1, pp: 225-237.

Biser, L C. (1983). Cooperative Farm Machinery Operation, 1970~1980. Performer: Agricultural Cooperative Service, Washington.

Carlo Russo; Dave Weatherspoon; Christopher Peterson; Massimo Sabbatini. (2000). Effects of Managers' Power on Capital Structure: A Study of Italian Agricultural Cooperatives. The international Food and Agribusiness Management Review, Vol.3, No.1, pp: 27-39.

Deng, J. L. (1989). Introduction to grey system theory. The Journal of Grey System, Vol.1, No.1, pp: 1-24.

Deng, J. L. (2005). Grey system basic method. Wuhan, China: Huazhong University of Blank Science and Technology Press.

Erchao Li; Minli Yang; Michael L. Cook. (2009). Agricultural Machinery Cooperatives in China: Origin, Development, and Innovation. 2009 American Society of Agricultural and Biological Engineers Annual International Meeting, Reno, Nevada, USA.

Georgeanne M. Artz; Gregory J. Colson; Roger Ginder. (2009). A Return of the Threshing Ring-Motivations, Benefits and Challenges of Machinery and Labor Sharing Arrangements. IOWA STATE UNIVERSITY, pp: 1-24.

Hodrick, R.J; Prescott, E. C. (1981). Post-war U.S. business cycles: An empirical investigation. Carnegie-Mellon University, pp: 451.

Ju, J.Y; Wang J.W. (2009). Prediction method for the operation level of agricultural mechanization in Heilongjiang Province. Transactions of CSAE, Vol.25, No.5, pp: 83-88.

Jin, Xiuman; Ma, Chenglin. (1995). Research on critical value of purchasing power for farm machinery. Transactions of the Chinese Society for Agricultural Machinery, Vol.26, No.1, pp: 5-10.

Jian Fu Zhang; Zhi Jun Wu; Ping Fa Feng; Ding Wen Yu. (2011). Evaluation systems and methods of enterprise information and its application. Expert Systems with Applications, No.38, pp: 8938-8948.

Li, F. X; Xu, Z; Zhao, J. (2008). Analysis of factors influencing flue-cures tobacco production in Henan province. ACTA TABACARIA SINICA, Vol.14, No.2, pp: 39-43.

Liu, F.J; Ren,A.S. (1997). Historical Evolution and Features of Farm Machinery Market[J]. Transactions of the CSAE, No.4, pp: 13-18.

Liu, Y.M; Tian, Z.H. (2008). Research on determinant factor of farm machinery equipment level in China. Journal of Agrotechnical Economics, No.6, pp: 73-78.

Lu, B. F; Zhang, Z. L; Zhu, M. et al. (2008). Discrimination and analysis of key influencing factors for agricultural mechanization development. Transactions of the CSAE, Vol.24, No.11, pp: 114-117.

M. Marrit Van den Berg; Huib Hengsdijk; Joost Wolf. (2007). The impact of increasing farm size and mechanization on rural income and rice production in Zhejiang province, China. AGRICULTURAL SYSTEMS, No.94, pp: 841-850.

Paolo Biondi; Danilo Monarca; Arduino Panaro. (1998). Simple Forecasting Models for Farm Trantor Demand in Italy, France and the United States. J. agric. Engng Res, No.71, pp: 25-35.

Peter Hill. (2005). Partnering for profit from machinery. Machinery, No.3, pp: 136-138.

Unakitan, G; Akdemir, B. (2007). Tractor demand projection in Turkey. BIOSYSTEMS ENGINEERING, No.97, pp: 19-25.

Wang, X. L; Ji, C. Y; Zhou, J; Jiang, L; Gu, B. X. (2010). Design and experiment of fruit and vegetable grasping system based on grey prediction control. Transactions of the CSAE, Vol. 26, No. 3, pp: 112-117.

Yang, M.L; Bai, R.P; Liu, M. (2005). Development of Agricultural Mechanization and Construction of Modern Agriculture. Transactions of Chinese Society for Agricultural Machinery, Vol.36, No.7, pp: 68-72.

Yang, Y. S; Liu, Z. Y; Sheng, G. H. (2005). Research on the promotion system of Chinese agricultural mechanization. Transactions of Chinese Society for Agricultural Machinery, Vol. 36, No.11, pp: 79-82.

Zhang, R; Gao H.W. (2009). Prediction Model of Agricultural Mechanization Level in China Based on GM(1,1). Transactions of Chinese Society for Agricultural Machinery, Vol.40, No.2, pp: 91-94.

Zhu, H. J; Tian, Z. H; Han L. J. (2007). Dependence of agricultural mechanization on financial investment. Transactions of the CSAE, Vol.23, No.3, pp: 273-278.

20

Development of Multi-Componential Decision Support System in Dynamically Changing Application Domain of Environment Protection

Dale Dzemydiene[1] and Saulius Maskeliunas[2]
[1]*Mykolas Romeris University*
[2]*Vilnius University Institute of Mathematics and Informatics*
Lithuania

1. Introduction

The integrated, working on-line systems create the possibilities to combine the information from all the participating data warehouses and giving an opportunity to extract useful information by the use of available functions and tools. The systems designed for the dynamically changing domains in solving sustainable development problems are rather complex. The multi–componential, interoperable structure of decision support system is needed to develop for the purposes of applicable assistance in dynamically changing environment. The decision support system (DSS) requires additional means for knowledge representation and for management of the on-line working evaluation of ecological situation of the region and decision making processes in web environment.

The realization of the intellectual, working on-line decision support systems requires the developing of the innovative web portal infrastructure with embedded intellectual components (Leibowitz, Megholugne, 2003; Zavadskas, *et al.* 2007; Bolloju, 2002, Baltrenas *et al.*, 2008; Dzemydiene *et al.,* 2008). Web services will provide all types of information for users with different responsibilities. Novel work organization methods, knowledge systems, modern information - communication technologies are of great significance in supporting sustainable development management problems in web environment. Web portal infrastructure is constructed for managing, saving and receiving environmental information in different levels of participator's responsibilities.

The aims of this chapter concern:

- The methods to collect appropriate interface structures (components, scenarios) for services control;
- Recognize the concrete situation from data warehouses using broad spectrum of statistical and operative analysis methods;
- Use the distributed information systems and means of wireless communication systems (i.e. programming components, protocols, sensors and devices).
- The developing of decision support models which allow integration of the different pieces of data available into coherent framework for intelligence purposes with knowledge and decision making components.
- To recognize meaningful alternatives and strategies for prevention measures during the problem-solving process by means of knowledge- based analysis.

The sustainable development management and administration are organized by envisaging of very important problems for:

- Development of industrial potential aims are forwarded for reduction of expenses, improvement of working conditions, diminishing of risk, and implementing of new activity extension opportunities;
- Understanding the abilities for analysis of risk and arising threats;
- Backgrounds of legal regulation of unsustainable factors;
- Assessment of situation and planning strategies and their creation;
- Understanding scientific achievements and technological progress;
- Analysis of sustainable development indicators and connection with possibilities to define the regional situation in many aspects.

The chapter is devoted to problems of developing of the component-based architecture of the integrated decision support system that afford the monitoring situation and intellectual analysis.

The principles of environment protection consist of multiple components and make up a totality of requirements that can be presented as standards of enterprise functioning, requirements of healthy human environment, permission for functioning, taxes for cause pollution, etc. These principles give rise to very serious problem of legal regulation, the significance of which must be analyzed in multi-dimensional complexity and relations with all kinds of legal system. The current situation recognized on data management and reporting in the water sector resulted in the following findings. The water types have strong division between groundwater, marine water and fresh (surface) water. A lot of data are available at many different locations, in different types of databases, with a high degree of dedication and specialization.

The aims of this work are to show the possibilities to integrate the distributed data-bases of water management into web portal understanding the requirements of social and pragmatic interoperability, and using physical technological possibilities to describe e-services for public administration. We use data exchange standards of extensible mark-up language (XML), simple object access protocol (SOAP) messaging and hypertext transfer transport protocol (HTTP), Web Service Definition Language (WSDL), universal description discovery and integration (UDDI) registry for realization of the main interface components of water management portal realization.

The core of this on-line working system is the water resource management information system (WRMIS) that allows the access to the information distributed in different data warehouses and to recognize the situation on the surface water quality in rivers and lakes, ground water, and point sources (emissions). A detailed description of the main components of monitoring and decision making is given. The advantages of web service-based solution are presented by solving problems based on services offered by information systems while estimating proposals in what ways to simulate situations, to make and intellectualized environment pollution estimation by an object. For the improvement of information system structure and services research work is pursued.

2. Problems of recognition of boundaries for analysis of regional sustainable development management domain in DSS

Environmental protection problems are very important now, and a significance of their solutions will rise in near future. An ecological situation of a region may be considered as

rapidly changing environment. In this subject area we face the problem of representing process development dynamism and rapid information change. In addition, the process poses rather a complex inherent inner structure and mechanisms of interaction between subsystems that are frequently expresses by temporal, geographic, and space dependence conditions.

The complexity of structures of processes, multiple subsystems with their own complex mechanism interacting as internal or external parts, time and space/geographical dependencies, a great volume of data acquired from the processes, and multiple-criteria decision making are essential features for the analysis and representation of such an application domain.

The solutions of risk management must be solved in accordance with deeper organizational efforts of interstate, inter-departmental and interregional cooperation in a new way of understanding the complexity and possibilities of unsustainable development. The conception of sustainable development includes the way to match two different and sometimes contradictory attitudes as follows: "development – progress – growth" and "stability – security – environment". Brundtland Commission has brought forward this dilemma. They were the first who had defined the objective of sustainable development: *"Sustainable development is the process that meets present requirements without comprising the ability of future generations to meet their own needs".*

Seeking to achieve sustainable development of the town, planners face with the underlying problems as follows:

- Municipalities don't possess land proprietarily rights;
- The majority of cities haven't prepared any strategic or general plans for town development that correspond to the conditions of market economy as well as the principles of sustainable development;
- New territories are being reclaimed for constructions, while social, technical and communicational infrastructure stay behind hopelessly;
- The majority of construction objects are being built within a very long time and this necessarily cause the high price for buildings;
- The work quality of city passenger transport is getting worse.

The general objective of sustainable development is to protect and improve quality of life. Therefore, both preventives need new innovative management methods which have a global influence on sustainability. People, however, appreciate many moments of environment that are not related with physical part of environment such as aesthetics, cultural environment, rural areas attainability and quietness as well. Again, inhabitants are concerned about very material things that are not environment at all. Material standards of living, social health and security, education opportunity, public health care, completeness of life, personal career possibilities, self-expression, community, culture, social life, recreation – all these things are treated as a part of quality of life.

To comprehend globalization phenomena, problems of economic development, trends of information and other technologies evolution, as well as the problems of animate nature, coordination of actions, social security and law are of the most significant trends for progressive development of regions (countries) under development. The social understanding of interoperability concerns intentions, responsibilities and consequences behind the expressed statements and commitments shared as results.

This research study is aimed to analyze activities of enterprises, institutions, and organizations according to some components of sustainable requirements dealing with the ecological sustainability, cleaner goods manufacturing, and economic growth.

The water management and water quality treatment is one of important problems related with environment protection, survival of variety of biological life cycles, and implies many requirements for sustainable management (Baltic 21 2003; Swanson *et al.* 2004, Dzemydiene *et al.* 2008).

The current situation recognized on data management and reporting in the water sector resulted: (1) Having different treatments of water types [groundwater, marine water, surface water, and others]; (2) A lot of data are available at many different locations, in different types of databases; (3) The amount of data to be stored and analyzed will dramatically increase in near future.

This carries out some difficulties to share data: no strong tradition for coordination and sharing of data and knowledge. The rising needs of water resources influence the sewage quality management issues.

In the light of these attitudes of sustainable development, researchers have raised the purpose to determine the possibilities of cities' development, at the same time evaluating various factors that influence the extents of investments in different towns for cleaner environment.

The activity of large enterprises, institutions, and organizations should be based on versatile responsibility of enterprises and stimulation of efficiency, paying ever more attention to the requirements of sustainable development and to the issues of environment protection: strategic and tactical planning and control, estimation of economic- social balance, application of information technologies and constant check systems, as well as to legal regulation effect.

Decision support is also related with the problems of estimation of general ecological situation of the given region, the indices of pollution provided in the project, the risk factors related to preservation of links that are of biological significance and time dependent, etc.

3. General solutions of component-based decision making system development

An approach integrates problem solving and model-based knowledge acquisition within an extensive model of different knowledge types, describing a changing information environment and decision-making processes in a DSS (Fig. 1). Some exceptional features of the changing environment, especially dynamic components, require additional means for knowledge representation, data verification and assurance of efficiently making decision processes.

An approach to modelling of problem solving expertise has been developed, based on the four layers of expert knowledge: (1) the domain layer contains definitional knowledge of the domain; (2) the inference layer describes the structure of reasoning and inference mechanisms that operate upon domain layer; (3) the task layer explicates knowledge about when a certain inference is to be made; (4) the strategic layer sets up the plan for achieving a goal, thereby controlling the reasoning in the task layer.

The expertise is knowledge in broadest interpretation of the term and includes factual knowledge of domain, problem solving strategies and methods, and learning strategies. Some components of the expertise model are distinguished in our decision support system as well: the fundamental knowledge that describes the application domain in static (semantic model of information structure) and dynamic (the imitational model of tasks and processes) perspective; the model of the problem solution strategy control level (e.g., plans,

diagnostic, strategies of correcting task fulfilment); the reasoning model that embraces stepwise decision support structure.

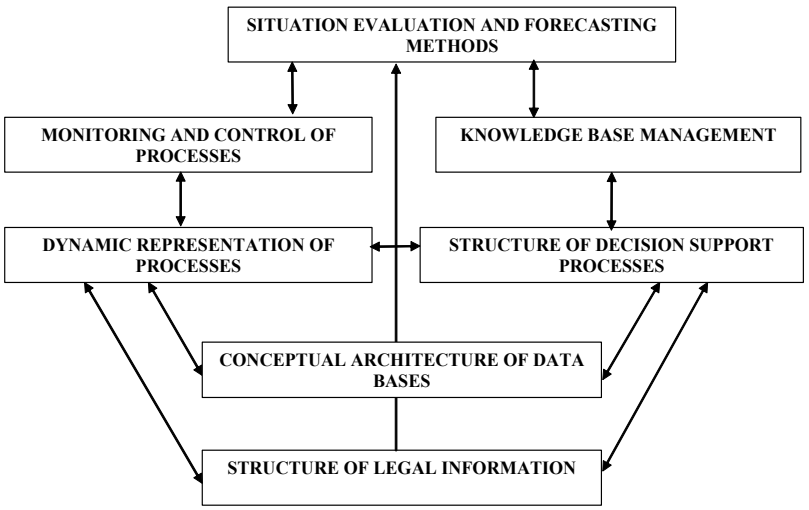


Fig. 1. General schema of the components of decision making system

The described decision support system helps in analyzing development processes of enterprises in accordance with water evaluation. We suggest that the means should be based on responsibility of enterprises and stimulation of efficiency, paying even more attention to the requirements of sustainable development.

The multiple objective decision support level deals with the analysis of information obtained from the all measurement points revealed in the dynamic sub-model of DSS.

The modelled system is regarded as direct mapping of real enterprise system and decisions can be based on decisive facts and follows rather deterministic rules.

Rapidly changing environment may include additional techniques for planning, operation control and decision support.

A new viewpoint and approaches are needed allowing us to concentrate the attention on the organizational aspects that ensure information for decision support. The technology for building such systems must provide methods for acquisition, structural representation of many types of knowledge taking into consideration the large, shared and distributed databases.

We understand the interoperability as "the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units" by recommendations of ISO 2382-01.01.47. Interoperability can be examined in different aspects of understanding of its framework, concerning physical, empirical, syntactical, semantic, pragmatic, and social layers.

The interoperability problems of distributed databases are important in the developing of the operatively working web services aimed for all sectors of managing. The following web services are designed for solving of tasks in water resource management and contamination evaluation sector with due attention to the international environment protection context.

The examples of development multi-componential decision support system are demonstrated on research investigations according to the requirements of European Union (EU) Water Framework Directive, Sustainable development Directives and EIONET ReportNet infrastructure. The main components of decision support system are analyzed by using different knowledge modelling and web service development techniques. The structure of water resource management information system (WRMIS) becomes the core of the decision support system in which web services are realized. The main components for evaluation of processes of contamination and water monitoring are represented by data warehouse structures.

The solutions to satisfy the interoperability requirements are demonstrated by architectural design of the system by integrating the distributed data warehouses and geographical information system means. The web services are based on common portal technology. The organizational and political arrangements require deeper and stronger participation activities by all members in reporting and understanding the importance of sustainable development problems and risk evaluation possibilities.

4. The architecture of component-based decision making system for evaluation of water contamination processes

The core of our decision support system and portal web service realizations is the information system – WRMS that realizes main interoperable functions of distributed information systems of water data storage. This water resource management information system (WRMIS) is developed under the project "Transposition of the EU Water Framework Directive in Lithuania" (Carl Bro as and AAPC 2003). The EU Water Framework Directive (Directive 2000/60/EC) and the future European environmental data reports repository ReportNet (Saarenmaa 2002) imply.

4.1 The constructions of the core system

Main principles of water resource management information system development: (1) a holistic approach, requiring integration of data and knowledge from different institutions and regions; (2) high data storing, analysis and reporting requirements; (3) assured interoperability based on XML Web Services, SOAP, HTTP protocols; (4) requirements for reporting in a format of GIS maps.

On the other hand, the detailed guidelines and software tools of ReportNet for Water sector data have not yet been finalized.

A focus area for the WRMIS prototype is to facilitate easy dataflow between the institutions and give access to data for relevant institutions and the public.

Most of the existing environmental information systems have evolved during a long period of time. Also, they stem from different traditions. Therefore most of the systems are found to be both heterogeneous and scattered - without much possibility of using data and information in an integrated manner. In the years to come it will be a necessity to combine data from many sources to better understand the environmental processes and to be able to make the required reporting. It is possible to overcome these barriers by creating an environmental portal.

Also it has been cleared up that the main general challenges to be met in developing such an environmental portal are as follows: (1) Ownership of data; (2) Telecommunication/Digital infrastructure (allowing sufficient Internet throughput); (3) Maintenance and development

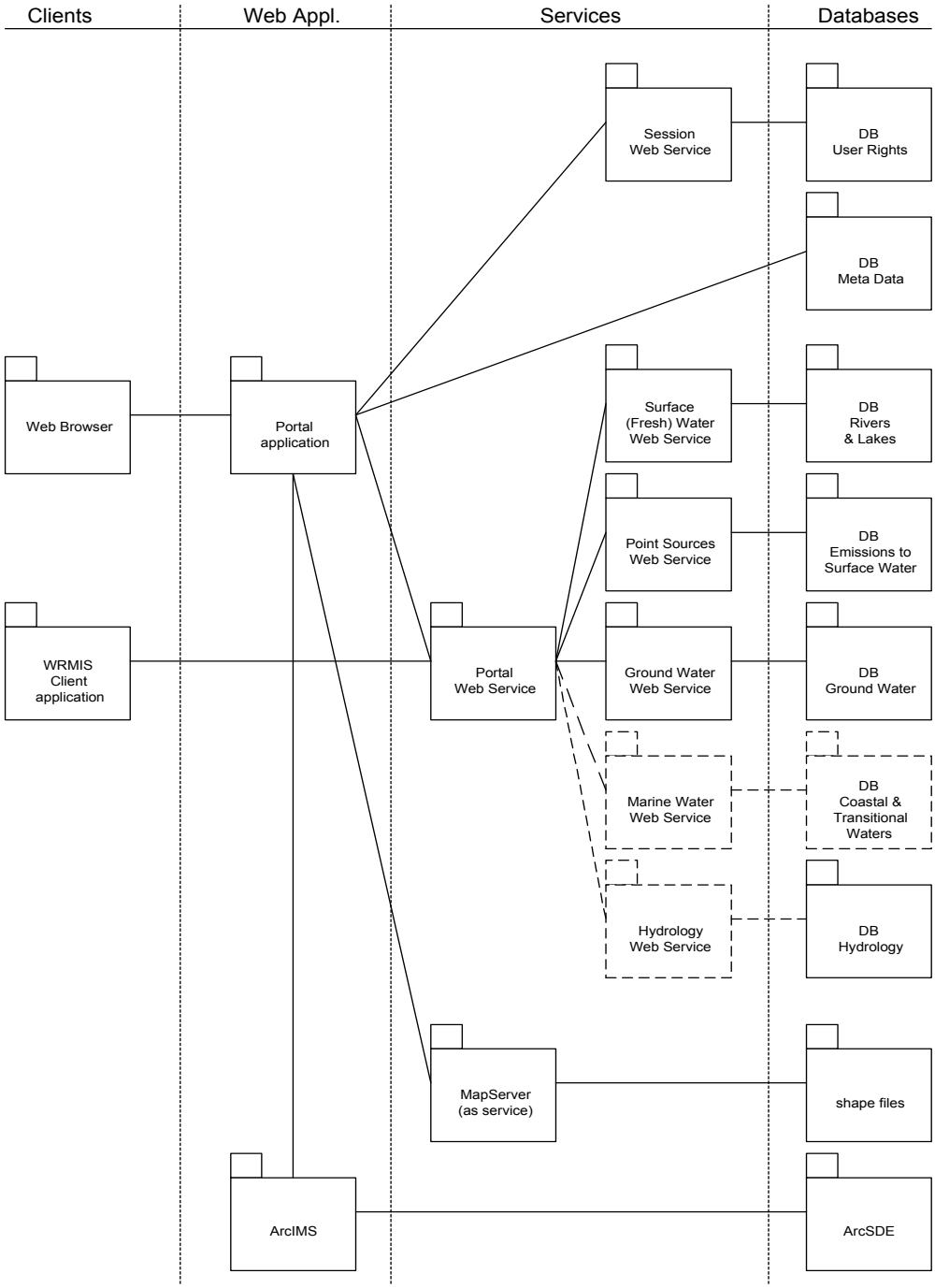


Fig. 2. The structure of interconnections of packages of the water resource management information system

(i.e., involvement of more institutions, regional departments of the Ministry of environment; integration with external data warehouses and Web Services - for reporting the EU, etc.): institutional setup, manpower; (4) Openness and a proper public participation.
The various types of data providers to the system require that some special software should be available for them. It can be achieved by implementing a Client/Server system based on services at the participating data warehouses.
The scheme of relationships between main components of the water resource management information system prototype, corresponding to the mentioned requirements of technical interoperability is presented in Fig. 2.
The databases and allocated tools have to be secured by restricted access.
A distributed database system, based on data warehouses (DW) and web services, was chosen because it improves both the quality of data and the value of the reports. It allows us to eliminate some of the boring and time consuming work associated with collection of data from different sources and bringing it on a format so that could be used in the common context. In addition, it has been decided that the system should be independent of which type of database provider is selected.
This is the case where the following tools are used: 4th-generation-language development environment suitable for a Rapid Application Development; the whole access to the databases should be performed by use of SQL; XML and its technologies should be used for data exchange and presentation.
Continually we have had evaluate the situation according to the sustainability in the given region. For this purposes we deal with the decision support, integrating main components for evaluation of environment contamination processes (Dzemydiene *et al.* 2008).
Decision-making aimed to on-line evaluation of the pollution processes of an enterprise deals with: the complexity of structures of processes; multiple subsystems with complex mechanism interacting of internal or external parts; time and space/geographical dependencies; great volume of data acquired from the processes; multi-criteria decision - making; causal, temporal relationships and interaction of processes.

4.2 Knowledge - based description of processes for contamination evaluation

The main purpose of the advisory system is to assist in environment protection control processes, creating of suspect profiles by giving computer-aided instructions, planning and situation recognition techniques. Use of advisory systems may also improve legal training, for example, providing environment protection agencies with advice on the type of information required by inspection to reduce the pollution activities. As an example, we consider the activities of enterprises, firms, and organizations, i.e. the main stationary objects, to estimate pollution of water bodies. In line with the object functioning nature the project reflects information on the activities pursued while the license defines the limit in which environment pollution is allowed, i.e. limits of pollutants cast are drawn. Besides, the objects must give reports according to their activities pursued and in line with statistical account ability forms.
The level of representation of dynamical aspects shows the dynamics of observable processes. The multiple objective decision making deals with the analysis of information obtained from the static sub-model taking into account all possible measurement points revealed in dynamic sub-model of such a system. Further actions, operations, etc. are determined through the mechanism of cooperation of agents that are working by using the temporal information registration window.

The complexity of environment research problems consists in the complexity of criteria and differences of attitudes.
The knowledge representation framework supports organizational principles of information in a static semantic model. The model of behavioral analysis of the target system shows the dynamics of observable processes. One of its characteristics is a need for a lot of data to properly model and verify these problems.

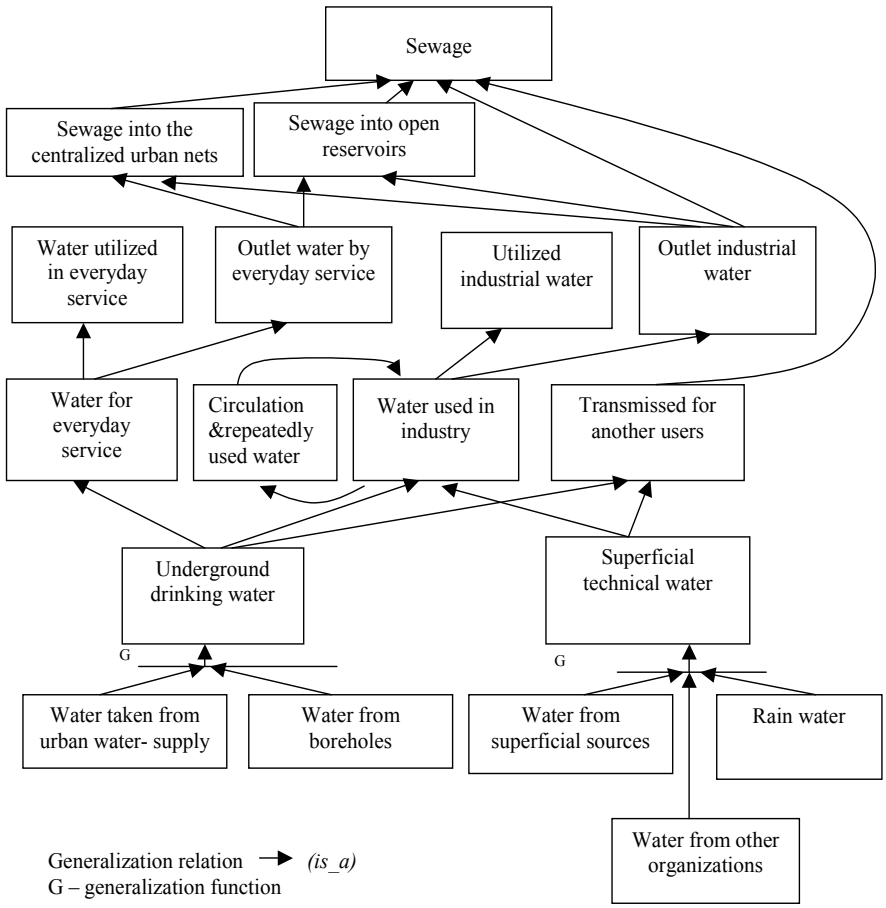


Fig. 3. An example of semantic representation of water distribution in a contamination object

Let X_{ext} denote the extension of a concept X_{con} with respect to its extensionality and X_{int} denote the extension of a concept X_{con} with respect to its intentionality. Then, the definition $(\forall x/X_{ext})(\exists pp/X_{int*})(\forall p/pp)p(x)$ is the intentional representation of a concept X_{con} .
Three types of abstractions of relationships between chosen concepts are used in constructing a semantic model: generalization, decomposition (aggregation), and transformation.
Generalization abstraction defines the type of intention "*concept - concept*": E_j *IS_A* E_i , where another generalizes one concept is defined by the set of another concepts.

Aggregation (decomposition) abstraction helps us to construct concept by others concepts depending on their decomposition or functional dependence: E_i *PART_OF* E_j

Transformation is very similar to that of the aggregation, except that it contains a calculation rule, which specifies how the values representing the occurrences of the defined concept are derived from the values representing the defining concepts.

More in details we consider the example of analysis of water resources and the pollution of sewage of the enterprise. The pollutants from the production are entering into the water in some types of cases (Fig. 3). The initial task is always data gathering, resulting in a set of observed findings.

A dynamically changing environment imposes time constraints. Many problems are to be solved simultaneously. The values of the observed parameters may change dynamically, depending on time and the events occurring. Solution of different problems is interfered with one another. For instance, the high concentration of harmful material thrown out into the air is related with the risk factors referring prevention of links that are of biological significance and time-dependent, etc. Another essential aspect of such an application domain is its spatial dimension. While in many other application domains the problems of study are within a very precise and, usually, narrow frameworks. For instance, the contamination problem of an enterprise (e.g. manufactory, firm, and plant) deals with spatially varying phenomena of unbounded limits.

4.3 Formal representation of processes by E-nets

The E-nets (Evaluation nets) are the extension of Petri nets, and were introduced by (Noe, Nutt 1973). The structure and behavioural logic of E-nets give new features in conceptual modelling and imitation of domain processes and decision-making processes (Dzemydiene, 2001).

Apart from time evaluation property, E-nets have a much more complex mechanism for description of transition work, some types of the basic transition structures, a detailing of various operations with token parameters. In addition to Petri nets, two different types of locations are introduced (peripheral and resolution locations).

The exceptional feature is the fact that the E-net transition can represent a sequence of smaller operations with transition parameters connected with the processes.

It is possible to consider the E-net as a relation on (E,M_0,Ξ,Q,Ψ), where E is a connected set of locations over a set of permissible transition schemes, E is denoted by a four-tupple: $E=(L,P,R,A)$, where L is a set of locations, P is the set of peripheral locations, R is a set of resolution locations, A is a finite, non-empty set of transition declarations; M_o is an initial marking of a net by tokens; $\Xi=\{\xi j\}$ is a set of token parameters; Q is a set of transition procedures; Ψ is a set of procedures of resolution locations.

The E-net transition is denoted as $a_i=(s_i, t(a_i),q_i)$, where s_i is a transition scheme, $t(a_i)$ is a transition time and q_i is a transition procedure.

The input locations L_i' of the transition correspond to the pre-conditions of the activity (represented by the transition in Fig. 4). The output locations L_i'' correspond to post-conditions of the activity.

In order to represent the dynamic aspects of complex processes and their control in changing environment it is impossible to restrict ourselves on the using only one temporal parameter $t(a_i)$ which describes the delaying of the activity, i.e. the duration of transition.

The complex rules of transition firing are specified in the procedures of resolution locations Ψ and express the rules of process determination.

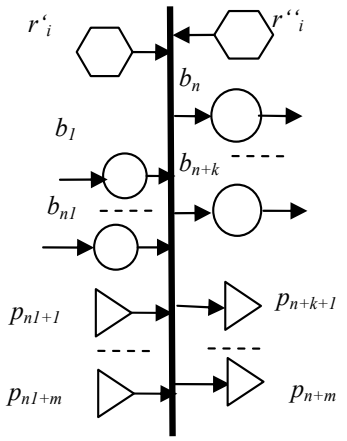


Fig. 4. The common transition schema of E-net representation

A concrete parameter of token obtain a concrete value according to it's identification, when the token is introduced into the location $b_j(\xi_k)$. Such a combination of locations with the tokens in them, the parameters of which obtain concrete values, describes a situation for process execution.

Such an understanding of the transition procedure enables us to introduce the time aspects into procedure of control of processes and determine operations with token parameters in time dimension.

The exceptional feature is the fact that the E-net transition can represent a sequence of smaller operations with transition parameters connected with the event/process. Operations are described in the transition procedure with these parameters $\Xi=\{\xi j\}$.

The E-nets support a top down design in graphical representation manner. The hierarchical construction of dynamic model is simplified by representing macro-transition and macro-location constructions.

The goal of the manufacture enterprises is to rationally develop an ecologically clean production. It means that enterprises (e.g. factories, plants, etc.) must guarantee the manufacture of products with minimal pollution of the environment and damage to nature, not exceeding the permissible standards.

Thus we have a problem of two objectives/goals:

G_1 – increase in capacity of production and profits of the enterprise under consideration;

G_2 – decrease in environmental pollution within the permissible limits.

The first criterion $\tilde{N}$ – the profit of production has to be maximized while the second criterion $\tilde{O}$ – the environmental pollution has to be minimized. The importance of the first criterion is determined by various economic industrial parameters. The maximization of its importance is of direct interest to and care of the producers. The main task of our decision support system is to maintain the importance of the second criterion $\tilde{O}$ within the permissible limits and thus act on the first criterion. The $\tilde{O}$ criterion is stipulated by three factors: water pollution, air pollution and contamination of harmful solid wastes: $\tilde{O}=\{W,A,H\}$.

More in details we consider the example of the analysis of water resources and the pollution of sewage of the enterprise. The pollutants from the production are entering into the water in some types of cases. Such cases we find out by the construction of E-net distribution processes of sewage in the enterprise (Fig. 5).

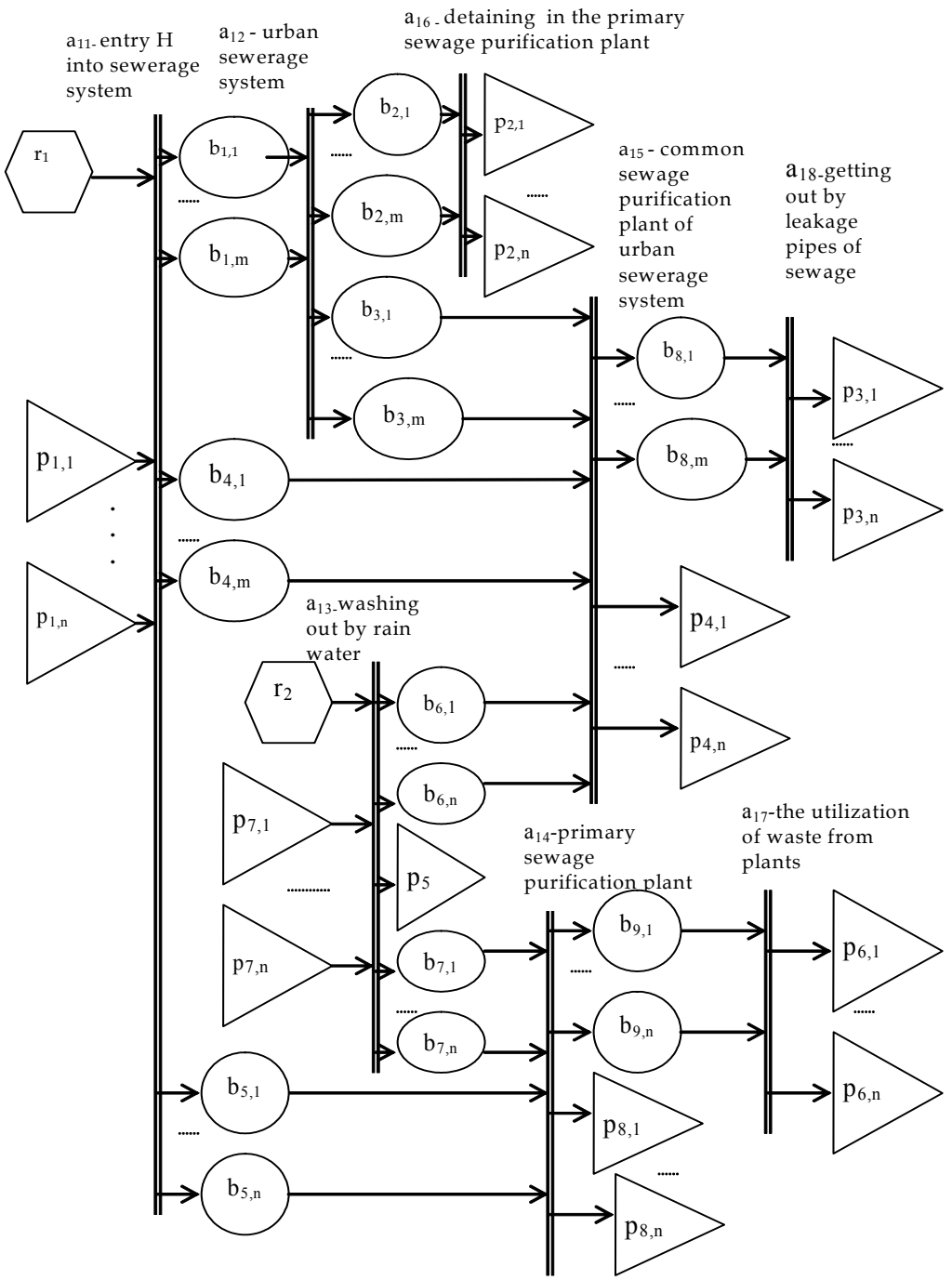


Fig. 5. The E-net of distribution processes of harmful materials in the water of enterprise

The harmful materials that are represented by peripheral locations of the E-net (Fig. 5) are very important for evaluation of water pollution of an enterprise:

$p_{1,1},\ldots,p_{1n}$ are materials included in water efflux;
$p_{2,1},\ldots,p_{2,n}$ are waste materials from the primary sewage purification plant;
$p_{3,1},\ldots,p_{3,n}$ are waste materials from the common sewage purification plant;
$p_{4,1},\ldots,p_{4,n}$ are materials entering into open reservoirs, that are not detained in the sewerage system of the enterprise;
$p_{5,1},\ldots,p_{5,n}$ are materials entering into open reservoirs, if there is no rainwater collection system;
$p_{6,1},\ldots,p_{6,n}$ are utilized wastes from primary purification plants;
$p_{7,1},\ldots,p_{7,n}$ are materials entering into rain water if they are stored openly in the territory of enterprise;
$p_{8,1},\ldots,p_{8,n}$ are materials entering into the external reservoir from the primary sewage purification plant.
The E-net structure, which describes the decision-making process, gives visually the parameters needed for control and the control structure relation with tasks and decisions.

5. Web portal structure for water management by requirements of interoperability of information systems

The portal has been developed as a web portal where the users can receive information related with WRMIS according to the water framework directive (WFD). It combines data from the participating data warehouses (DWs) and gives the users an opportunity to extract useful information by use of the functions and tools available.
The portal has been developed by use of Active Server Pages (ASP - server-side script engine for dynamically-generated web pages0, and it communicates with the DWs by use of SOAP/XML (see an example of ASP code for generating SOAP calls in Fig. 6).

```
Dim xmlDocRoot, xmlDocChildren
Set XMLClient = Server.CreateObject("MSXML2.ServerXMLHTTP")
SoapText=SOAPCallBegin+"<SOAPSDK1:GetSites xmlns:SOAPSDK1=""urn:SfinxIntf-ISfinx""">"+_
    "<ASessionId>" + CINETSessionID + "</ASessionId>" + _
    "<ARegionNo>" + ARegionNo + "</ARegionNo>" + _
    "<ASiteId>" + ASiteNr + "</ASiteId>" + _
    "<AStartDate>" + "1995-12-30" + "</AStartDate>" + _
    "</SOAPSDK1:GetSites" + SOAPCallEnd
SoapServer = SOAPServerURL
Call XMLClient.open("POST",SoapServer,"False")
Call XMLClient.setRequestHeader("Man", "POST"+" "+SoapServer+" HTTP/1.1")
Call XMLClient.setRequestHeader("MessageType", "CALL")
Call XMLClient.setRequestHeader("Content-Type", "text/xml")
Call XMLClient.send(SoapText)
Set ObjXML = XMLClient.responseXML
Set xmlDocRoot = ObjXML.documentElement
Set xmlDocChildren = xmlDocRoot.childNodes
Set xmlDoc = Server.CreateObject("Microsoft.XMLDOM")
xmlDoc.Load(ObjXML)
Set objXSL = Server.CreateObject("Microsoft.XMLDOM")
objXSL.async = false
styleSheet = StylesheetDir & "\siteinfo.xsl"
objXSL.load(styleSheet)
```

Fig. 6. Example of SOAP call to "getSites" by use of ASP

The portal gives access to information on: surface water quality, ground water, and point sources (emissions).
The portal includes services where the users can have useful information related to the WFD and the WRMIS, including links, documents, and news.
A login procedure has been established on the WRMIS portal to secure that only registered users can have access to the available information. The WRMIS system administrator has special rights to the system, that give access to maintain the system by updating the news, links, documents, meta database, and the system administrator is responsible for granting rights to the users of the system. The list of portal Web Service operations is presented in Table 1.

Operation	Description
ListOfRegions	Returns a list of region IDs and names
ListOfSites	Returns a list of site IDs and names
GetSiteInfo	Returns site information, such as name, coordinates, distance to river mouth, etc.
GetSamples	Returns a list of sample dates for the specific site
GetAnnualMeans	Returns annual means for a specific monitoring site in a specific period
GetAnalysisResults	Returns raw data for a specific monitoring site and period
ListOfPS	Returns list of point sources
GetPSInfo	Returns main information for a specific point source
GetPSOutlets	Returns raw data for a specific point source and period

Table 1. The main operations of web services of the developing portal

The UDDI registry of WRMIS Web Services is developed for internal purposes, having in mind that exposing it freely on the Internet increases the security risks (Data Junction 2003).

5.1 Surface water database (DW1)

A data warehouse has been established based on data from the "Water monitoring in rivers and lakes (VANMON)" database. Database structures similar to the VANMON database have been created in Oracle, and an interface by use of a database SOAP server has been established on top of the database. The database server provides SOAP-based Web Services with basic functions for using and maintaining the database, e.g.: selection from, update of, and appending to the WRMIS VANMON database.
By connecting to the WRMIS portal and selecting a Surface Water item in the menu, the user is able to retrieve data from a specific river monitoring station. User has a possibility to select a site and get all the analysis results stored in a table form, and to save the retrieved data as an Excel spreadsheet.
The dropdown box with regions is the result of a SOAP call to the Surface water data warehouse. When the user selects a region in the dropdown box, or when the user clicks on a region in the map, a new map and an additional drop down box with sites appears. This map can be zoomed to the selected region. In the map the site and river layers are visible, and the site layer is click-able.
The surface water Web Service has the next operations: Init, Meta Tables, Meta Fields, Reload Meta Data, Add To Table, Delete From Table, Update In Table, Select From Table, Select From Table(2), Joined Select, Replace Lookup Select, Union Select, and Get Object.

5.2 DW1 client application

For the WRMIS VANMON database a special client application has been developed, that enables the users to update the central WRMIS VANMON database by use of the Internet. The application makes it possible for the staffs both in the Regional departments and in offices of EPD to work with the common data.

The application can be downloaded from the WRMIS portal and installed on a local PC. The use of this application is managed by the WRMIS VANMON database administrator, who grants user rights to all users.

The client application, the portal SOAP server, and the database SOAP servers have all been developed by use of Delphi. The server applications run on the Internet Information Server, while the client application can run in a Windows environment on the local PCs.

5.3 Point source database (DW2)

A second data warehouse similar to the surface water data warehouse was established on top of the point source (emissions) database. The system gives access to information on outlets from point sources. The database was made up by redesigning and integrating separate annual FoxPro databases into one common database.

The functionality of the Point Source database and the corresponding Web Service is rather similar to that of the Surface Water database.

5.4.Ground water database (DW3)

The third data warehouse has been found in the Geological Survey of Lithuania (GS). The WRMIS staff in GS has elaborated this data warehouse in the Java environment. Web Services that give access to information stored in the GS Oracle Ground water database have been developed and made available on the Internet.

The technology used: (1) Programming language [lgt.* classes]; Java JSDK 1.4 API; (2) SOAP server [SOAP Servlet], Apache SOAP v2.3.1; (3) Servlet Container [Web server], Apache Tomcat 4.04; (4) Database [geological dataset], Oracle8 Enterprise Edition Release 8.0.6.2.0; (5) Link to Database, Oracle JDBC Thin JDBC Driver; (6) Server Operating System, SunOS 5.8 [Solaris 8] Generic_108528-17 sun4u sparc SUNW, UltraAX-i2; (7) Server Platform [hardware], Sun Fire V120 Server, 650 MHz UltraSPARC-IIi.

The WRMIS prototype has a wide use of digital maps in the GUI both for presentation and administration purposes. All digital maps are stored either in ArcSDE, or in shape files. The system uses maps for selecting data. Two map servers have been developed, making it possible for the users to receive information in a map and to compose their own maps on request. The first solution has been developed by means of the MapServer (hosted in EPA) and shape files.

The MapServer is an open source development environment[1]; MapServer functionality is easy to integrate in Internet-based GIS applications. Another solution has been achieved by use of ArcIMS and ArcSDE (hosted in the Ministry of Environment).

A meta-database with information on the digital maps used in the Map tools has been formed. The meta-database holds information described as mandatory in the ISO 19115 standard[2].

---

[1] http://mapserver.gis.umn.edu/

[2] http://www.isotc211.org/scope.htm#19115 , http://www.isotc211.org/publications.htm

The system is ready for data interchange with external Web Services. Also, the system can be easily improved to share information on digital maps with external systems by implementing, using and supporting the OpenGIS Web Map Server (WMS) and Web Feature Service (WFS). Both Mapserver and ArcIMS can be configured to support WMS.
A large amount of measurement points at different time and conditions causing overlapping and conflict between different observations, and the inaccuracy of measurements and reports.

6. Conclusions

Consequently, intelligence programs have been developed, leading to recognition of a field of activity called contamination analysis, which has been described as the identification of and the provision of insight into the relationship between environment data and other potentially relevant data with the view to specialist experts. A key part of this approach enforcement is to understand those activities, through the development and use of methods, models and tools for collecting and then interpreting the large volume of data available in real time for environment protection investigation. Some issues for qualitative information representation including statistical analysis are considered.
The water resource management information system proves that it is possible to establish a Web portal that provides the users with information, on request, in a system based on decentralised data sources. This project has also proved that it is possible to integrate information from different institutions, based on different technology, into a common information system by means of the Internet, XML and Web Service technologies, and by bridging to the well known existing systems.
The WRMIS system architecture is designed both to serve a centralised and a decentralised solution. It is therefore possible to configure a system so that the database servers were placed at different geographical locations as well as at the same place. A centralized solution with all the servers placed in one centre can be maintained by a highly skilled staff that can both maintain the hardware, and the software. Such a centre ought to have powerful equipment and Internet connections and the uptime for services should be close to 100%. The disadvantages are that data, in many cases, are at the hands of people that have none or but little knowledge on the use of data. There is therefore some risk that the staffs have vague motivations for the system performance and further development.
Another solution can be based on decentralized data warehouses, where different topic centers are responsible for the maintenance of their own data warehouse. This requires for competent manpower able to have the servers running as well as to administrate and maintain the software and databases at each data warehouse.

7. Acknowledgment

We would like to acknowledge the developers from Denmark and especially engineer Kim Jacobsen from Carl Bro regarding the possibilities to participate in the processes of preparing the project "Transposition of the EU Water Framework Directive in Lithuania" and realization of water resource management information system (WRMIS), and Ministry of Environment of the Republic of Lithuania.

8. References

Analysis of Summary of Results of Environment Protection Control. 2006. Data of State Environment Protection Inspection. http://vaai.am.lt/VI/index.php#a/458

Baltic 21, 2003, Report 2000-2002: Towards Sustainable Development in the Baltic Sea Region. Baltic 21. Series No. 1. Poland, Drukarnia. MISIURO.

Baltrėnas, P.; Morkūnienė, J.; Vaitiekūnas P. (2008). Matematical Simulation of Solid Particle Dispersion in the Air of Vilnius City, Journal of Environment Engineering and Landscape Management 16(1): 15-22.

Brown, A.; Johnston, S.; Kelly, K. (2002). Using Service-Oriented Architecture and Component-Based Development to Build Web Service Applications. IBM Rational Software White Paper, http://www.rational.com/media/whitepapers/TP031.pdf

Carl Bro as and AAPC, 2003. Transposition of the EU Water Framework Directive and Elaboration of a National Strategy for the Management of Water Resources in Lithuania. Project of DANCEE (Danish Co-operation for the Environment in Eastern Europe) programme, http://gis.am.lt/wfd/

Data Junction. (2002). Web Services: The New Generation of Application Integration. White Paper. http://www.datajunction.com

Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for the Community action in the field of water policy. *Official Journal*, OJ L 327, 22 December 2000. Available from
http://europa.eu.int/comm/environment/water/water-Framework/index_en.html

Dzemydiene D. (2001). Temporal Information Management and Decision Support for Predictive Control of Environment Contamination Processes. In J. Eder, A. Caplinskas (Eds.) Advances in Databases and Information Systems. Proc. of Fifth East-European Conference. Vilnius, p. 158-172.

Dzemydiene D. 2001. An Approach to Modelling Expertise in the Environment Pollution Evaluation System. In Databases and Information Systems. J. Barzdins, A. Caplinskas (Eds.) Kluwer Academic Publishers. Dordrecht/Boston/London, p. 209-220.

Dzemydiene D., Naujikiene R. 2003. Structural Analysis of Legal Environment Protection Information in the Estimation of Pollution Factors. In: H. Pranevicius, E. Zavadskas, and B. Rapp (Eds.), Proc. of International Conference "Modelling and Simulation of Business Systems". KUT Technologija press, p. 324- 328.

Dzemydienė D., Maskeliūnas S., Jacobsen K. 2008. Sustainable management of water resources based on web services and distributed data warehouses// Technological and Economic Development of Economy. Baltic Journal on Sustainability. Vilnius: Technika, Vol. 14, No. 1, p. 38-50.

Egyedi T.M. (2007). Standard-compliant, but incompatible? Computer Standards & Interfaces. Vol. 29. Issue 6, p. 605-613.

Environment 2010: Our future, our choice. The Sixth EU Environment Action Programmer 2001-2010. Luxembourg. 2001. Available from (http://www.environment.com).

Jacobsen, K.; Maskeliunas S. (2003). Web service-based information system of water resource management. In: Jean Claude Rault (Eds.), ICSSEA 2003, 16th International Conference "Software & Systems Engineering and Their Applications", Proceedings. 5. Centre pour la Maitrise des Systemes et du Logiciel, CNAM, Paris, France, p. 1/10-10/10.

Karnegy B. W, Filion Y. R. (2005) Linking Health Concepts in the Assessment and Evaluation of Water Distribution Systems // *Bulletin of Science, Technology & Society*. Vol. 25, No. 3, p. 247-253.

Komatsoulis, G.A.; Warzel, D.B.; Hartel, F.W.; Shanbhag, K. ; Chilukuri, R. ; Fragoso, G. ; Coronado, S.; Reeves, D.M.; Hadfield, J.B.; Ludet, C.; Covitz, P.A. (2008). caCORE version3: Implementation of a model driven, service-oriented architecture for semantic interoperability // *Journal of Biomedical Informatics.* Vol. 41, Issue 1, p. 106-123.

Legal information system LITLEX. Available from http://www.litlex.lt/portal/ml/start.asp?lang=eng. 2006.

Mysiak J., Giupponi C., Rosato P. (2005) Towards the Development of a Decision Support System for Water Resource Management // *Environmental Modelling & Software.* Vol. 20, Issue 2. 2005, p. 203-214.

Moavenzadeh F., Hanaki K., Baccini P. (Eds.) (2002). *Future Cities: Dynamics and Sustainability*. Series: Alliance for Global Sustainability Book series, Vol.1, p.-248

Noe, J.D.; Nutt, G.J. (1973). Macro E-nets for representation of parallel systems // *IEEE Transactions on Computers,* C-22(5), 1973, p. 718-727.

Rajasekaram V, Nandalal K.D.W. (2005) Decision Support System for Reservoir Water Management Conflict Resolution // Journal of Water Resources Planning and Management-ASCE. Vol. 131. Issue 6, p. 410-419.

Richter B. D., Mathews R., Harrison D.L., Wigington R. (2003) Ecologically Sustainable Water Management: Managing River Flows for Ecological Integrity // Ecological Applications, Vol. 13, Issue 1, p. 206-224.

Liebowitz J.; Megholughe I. (2003). Asset of Frameworks to Aid the Project Manager in Conceptualizing and Implementing Knowledge Management Initiatives // International Journal of Project Management. 21 (5): 189-198.

Swanson D.A., Pintér L., Bregha F., Volkery A., Jacob K. (2004). National Strategies for Sustainable Development: Challenges, Approaches and Innovations in Strategic and Co-ordination Actions. IISD.

Saarenmaa H. 2002. Reportnet: Development of Common Tools and a Shared Information Infrastructure for the European Environmental Information System. NFP/EIONET Group Meeting papers, 2nd edition, http://www.eionet.eu.int/EIONET_Tech/sii

Zavadskas E.K., Kazlauskas A., Maciunas E. Vainiunas P., Marsalka A. 2007. Real Estate's Market Value and a Pollution and Health Effects Analysis Decision Support System// Yuhua Luo (Ed.) Cooperative Design, Visualization, and Engineering, 4th International Conference, Lecture Notes in Computer Science 4674. Springer. CDVE, 191-200.

21

Decision Support Systems in Water Resources Planning and Management: Stakeholder Participation and the Sustainable Path to Science-Based Decision Making

Aleix Serrat-Capdevila, Juan B. Valdes and Hoshin V. Gupta
International Center for Integrated Water Resources Management ICIWaRM-UNESCO
Department of Hydrology and Water Resources
The University of Arizona
USA

1. Introduction

This chapter will focus on decision support systems (DSS) as they relate to water resources management and planning. Water is a resource that touches and is interwoven with numerous human activities as well as the environment we live in. Its availability and beneficial use depend on the timing and manner of its arrival (rainfall intensity, rain or snow, duration, frequency), the physical setting of the region (climate and weather, topography, geology), the engineering structures in place, the environmental constraints (existing ecosystems), the legal regulatory context and institutional policies. In most contexts, cultural values and preferences are also very important. To make good decisions, it is clear that a detailed understanding of how the system works and behaves is necessary. It is equally important to understand the implications of these decisions - what consequences are likely to ripple through the interwoven system, and what parties will be affected as a result of a particular set of actions? Understanding the coupled human and physical system is essential.

In addition to looking at the evolution of decision support tools and methods for water resources management (Section 2), this chapter focuses on how integrative science and multi-resolution models provide the basis for a decision support system (Section 3), on the overall setting of the decision making process and ways in which a DSS for water resources should be developed (Section 4). We make the argument that for a DSS to be successful and informative, the process by which it is developed will be as important, or even more so, than the finished decision support tool itself. A description of successful participatory planning approaches and collaborative modeling methods is presented, as well as a comparison of several case studies. Section 5 presents an overview on how to deal with uncertainty. We present our vision to merge adaptive management, integrative modeling and stakeholder participation to face the water management challenges of the arriving future. A synthesis and future challenges are presented in the last section.

2. Background: Water resources and DSS

Traditionally, decision support systems in water resources management have been characterized by limited decision-making scope. These decision support systems have typically been based on black-box optimization models, understandable only by technical people, and developed for very specific purposes (such as reservoir and infrastructure operations, engineering designs, etc). In general, such DSSs drew from a broad set of tools aimed at informing and supporting decision making, including a) GIS and other visualization tools to better 'read' and understand data, b) tools to help understand costs and effects of construction alternatives depending on design specifications, c) operating tables or models indicating actions to perform given a set of different coexisting constraints, and d) simulations to understand consequences of different operating policies or management alternatives, among many others.

In the US, there has been a move to consider these aspects since the 19th century, but the focus has been mostly on economic impact. For example, the 1936 Flood Control Act required only that the benefit–cost analysis be positive for a plan to be deemed feasible, and subsequent documents consolidated the concept of "*contribution to national income*" as the preeminent water resources planning objective (Loucks et al, 1981). Consequently, economic objectives – measured through benefit-cost analysis – have dominated water resources planning in the United States and worldwide, during much of the past century.

During the Harvard Water Program (1955-60), academicians and senior federal and state agency employees worked together on research and training for water resource systems design and planning. One of their principal goals was to "*improve the methodology of system design in such a way that it will meet any reasonable economic objectives within reasonable institutional constraints*". In other words, they developed tools and methods which, given a planning objective, would determine what set of structural measures, operating procedures, and water allocations ('*level of development for different water uses*') would best achieve the objective. They developed the use of multi-objective optimization methods, and proposed objective functions for economic development that could also account for other important aspects. The seminal book that came out of this program (Maass et al, 1962) describes its major accomplishments. Many of its methods remain in current use today for evaluating and ranking design alternatives based on economic efficiency.

In an attempt to address some of the difficulties of assigning economic values to the broad range of possible water resources planning objectives, the US federal government adopted (in 1973) the *Principles and Standards* of the Water Resources Council (revised in 1979) making environmental quality equally important to economic development as a planning objective. Gradually, there was a transition in which benefit-cost analysis went from being the primary objective to becoming a constraint required to ensure the economic soundness of a plan, among and equal to other considerations (Loucks et al, 1981).

However, even when planners and decision-makers acknowledged the need to account for other factors beyond benefit-cost and other quantitative analysis, the planning process was almost always engineered through the lens of computer modeling, as evidenced in the following citations: "*there are two basic approaches for solving planning models: simulation and optimization*" (Loucks et al., 1981, p.21); and "*The principal way [...] to identify, predict and evaluate the impacts of alternative plans or policies is through the development and use of mathematical models*" (Loucks and da Costa, 1991, p.3).

A good representation of the state of the art in 1990 is given by the proceedings of an international workshop on DSS for water resources research and management (Loucks and de Costa, 1991). It is interesting to note that the majority of its 24 articles are focused on software structure (pre and post-processing, databases, numerical models) user interfaces and visualization of results. Beyond mentioning the necessity for dialogue with the 'model client', very few articles referred to interactions with the end users, much less the stakeholders that were to be affected by the decisions. A notable exception is an article about USACE Hydrologic Engineering Center (HEC) software products, which attributes widespread use of its products mainly to interactions with the users: it describes a problem-driven research approach that listens to the users and tries to understand their specific needs, its program to train users, and the need for long term support in model implementation and analysis.

With few exceptions, therefore, models were developed mostly in support of the tasks to be performed by planners, managers and decision-makers, and were detached or disengaged from the challenges of being a decision-maker operating within the constraints of their constituencies and their part in the decision-making process. Not surprisingly, these prescriptive models were developed by engineers and technocrats, often viewed as the only source of trusted information, and with little or no stakeholder input (Cardwell et al. 2010). Historically this has caused difficulties in the implementation of decisions, resulting in lower than expected model usefulness, and low rates of project success.

3. Integrative science and models

The traditional approaches – with their optimization algorithms and their objective functions – were unable to successfully include into their computations the variety of important factors that are important to decision makers, and in ways that are transparent to the public. Their engineering-focused methods were unable to properly assign numbers to societal preferences and environmental values. Further, they were unable to reflect the possibility of solutions involving negotiated trade-off in a transparent way. There was no mechanism for representing the values of intangible assets, essential but invaluable variables, or the long term impacts on the resources of the commons (air quality, riparian impacts, land cover and landscape values, etc.).

During the past two decades the need for holistic approaches and cross-disciplinary teams that can address complex interactions at the basin scale and can evaluate alternative futures has become increasingly more evident. Integrated Water Resources Management (IWRM) has emerged as the new paradigm for decision-making in relation to water. This approach adopts the basin scale as the natural unit enabling water issues to be considered both in their broader context and through the more focused lenses of economic efficiency, social equity and environmental sustainability. This progression towards a holistic view of water resources research and decision-making has become reflected in new initiatives and programs within funding and donor agencies, sometimes making cross-disciplinary collaboration a basic requirement.

The need to handle information from diverse physical and social datasets, and to develop holistic and integrative decision support systems, has given rise to a new type of modeling tool in water resources planning: namely '*system dynamics modeling*'. Initially developed at MIT in the late 1960's (Forrester, 1968) for economic and business applications, system dynamics platforms facilitate flexible representations of the relevant behaviors from each

component of the system, and the incorporation of feedback loops. The book *Limits to Growth* is a good example of this, as it was based on a system dynamics simulation of the earth's population growth and resource use (Meadows et al., 1972). By design, they allow the decision-makers to see the entire forest through the trees, instead of getting lost in the details of each field and its specialized models. By including only what is important from each component of the system, they make it easier to represent and gain understanding of interactions among the different components of the system.

Everyone will agree this is a complex task. If a functional holistic and integrative model is to be developed to support decision-making, it is likely that this model will draw from findings and information from models specific to each system component. Regarding natural processes from the physical system, it will benefit from more spatially explicit and detailed models. Such a multi-resolution integrated modeling approach may be essential to face multi-disciplinary research and management challenges. Models of different resolutions will allow representation of different aspects of the problem and can be geared to answer different research questions and inform different sets of decisions (Liu et al 2008). For example, high resolution models (~100m grid cells) can represent in great detail the processes in the physical environment such as the land-atmosphere partitioning of water, the role of vegetation, the interactions between surface and groundwater hydrology or the dynamics of the saline wedge in coastal systems. These fine resolution models provide the state-of-the-art scientific understanding of the physical system. They allow us to extract the key aspects regarding the functioning of the physical system to be included in the medium resolution models (~1-12 km). Models at medium resolution combine (1) a less complex but accurate representation of the natural environment and (2) the human interventions on the environment such as land use management, engineering infrastructure and its operation in terms of intercepting and moving water within the basin. These medium resolution models allow us to represent the water allocations and re-distribution within the system and bridge the gap with the coarse resolution models. The higher level (coarse) models are the best attempt to represent the socio-economic and institutional aspects of water management over a simplified representation of the natural and engineered system, with a resolution at the scale of the sub-watershed.

Besides being able to answer different kinds of research questions, the benefit of a multiple resolution modeling approach is that information and findings can be transferred - and used to fine-tune - across models. While information regarding natural processes, impacts and feedbacks in the natural system can be up-scaled from the fine resolution to higher-level models, the behaviors and policies from the socio-economic and institutional models can be used to drive lower resolution models and assess impacts on the natural system. This approach has been formulated and described in detail by Wagener et al (2005) and Liu et al (2008) based on the experience of the NSF Science and Technology Center SAHRA (Sustainability of semi-Arid Hydrology and Riparian Areas) in conducting integrated multidisciplinary research addressing water management challenges in the US southwest.

Ultimately, planners and decision-makers are likely to use the modeling tools that simulate the overall behavior of the basin with a simplified but still accurate representation of all its components. Such a tool will represent the relevant behaviors of the system to answer their specific management questions. Because it draws from the findings of more complex models, this DSS model will be more computationally efficient, allowing numerous model runs in a short time. Roach and Tidwell (2009) and Kang and Lansey (2011) are excellent examples. The possibility of comparing simulations of different management options and decision alternatives through a user interface in a short time span makes system dynamics a

very user-friendly DSS tool for decision-makers and the public. Indeed, system dynamics DSS have recently been used to support basin scale mid and long range planning, and management (Tidwell et al. 2004; Yalcin and Lansey. 2004; Kang and Lansey, 2011). Two integrative system dynamics case studies to support planning and decision making can be found in the Middle Rio Grande in New Mexico, and in The Upper San Pedro River in Arizona. Both basins face severe water management challenges and need to find solutions to balance existing human and environmental demands with existing water supply.

In the Upper San Pedro, where human extractions from the basin aquifer threaten a Riparian Natural Conservation Area (SPRNCA), a mandate was passed by the U.S. Congress summoning the agencies and stakeholders in the basin to find a sustainable solution by 2011. In addition to the mandate, the possibility of the main economic motor of the basin (Fort Huachuca, a military base) being moved to another region if the water sustainability problems weren't solved, was a strong incentive to act. The development of the DSS by faculty and students at The University of Arizona benefitted from strong science contributions and the collaboration with numerous local stakeholders and agencies conducting research in the basin. The model allows users to select different packages of water conservation measures to be implemented through time and space in the basin. After each simulation, estimates are obtained regarding the impacts and improvements of the selected measures on the water budget, groundwater levels in key locations, and other parameters such as the costs of implementing such measures. The model is able to represent impacts on the groundwater system and the riparian area that depend on socio-economic profile for the basin and on the water conservation measures applied by the user. Linearized relationships between groundwater pumping and aquifer water levels were derived from a state-of-the-art groundwater model of the basin – a detailed physical model with higher spatial resolution – and included in the DSS for computational efficiency. The interested reader will find detailed information of the development of the San Pedro basin DSS model in Yalcin and Lansey (2004) and in Kang and Lansey (2011).

The Middle Rio Grande DSS model, developed by Sandia National Labs in collaboration with The University of Arizona, also benefitted from multi-resolution modeling. However the inclusion of information from detailed physical models was done differently. From a detailed high-resolution hydrologic model of the basin, a simplified one was derived lumping cells with similar attributes and hydraulic behavior. From a complex model with more than 100,000 cells, a simple one was produced with only 51 compartments (~cells) and sufficient accuracy to capture the overall behavior of the complex model, thus providing estimates at a level useful for policy analysis (Roach and Tidwell, 2009). As expressed by Passell et al (2003): "*this systems-level planning model draws heavily on the inferences and results of many other more sophisticated models focused on particular aspects of the basin*".

4. The sustainable path bridging science and decision-making

In general, scientists, academicians and some practitioners are convinced that numerical models are indeed a good tool to support decision-making, but the reality is that the adoption of modeling tools by policy and decision-makers is not standard practice. The main reason behind this fact is that, being extremely busy; managers, policy-makers and elected representatives are unlikely to use a model or tool they are unfamiliar with, regardless of how good it may be. Further, they will generally not use such models if they don't feel they understand how the models have been developed, and in what ways the

model has been designed to help them make informed decisions. Further, any decisions based on information provided by the models will not be considered sufficiently trustworthy if the models are perceived by the stakeholders as a) not being transparent, and/or b) if they are not convinced the model addresses their views and concerns, and/or c) their input has not been requested or integrated into the development of the model.

So, how can we merge the science, data and models with decision-making at different levels of operations, policy and governance, in a sustainable way over the long term? For all the integrative science described in the previous chapter to be perceived as credible, relevant and transparent (Liu et al 2008; Gupta et al 2011) – thus acceptable to inform and guide decision-making in the public eye – one key factor is essential: stakeholder participation through science-fed collaborative processes. In a participatory planning process, technical models used to support planning and decision-making are developed collaboratively. That is, decisions embedded in models are a product of agreement – sometimes after extensive discussion – between scientists and stakeholders during periodic meetings. Such model development forces the individuals involved to focus their communication on important issues, ranging from processes and features represented in the model, to assumptions, conservation measures, alternative scenarios, etc. This process provides an excellent setting for ongoing simultaneous discussions about specific issues, being key to a better understanding of the overall behavior of a system, the nature of certain problems and potential solutions. Importantly, the participants educate each other, and a better overall understanding is gained at many levels. First, it helps stakeholders understand the physical system, and in particular the spatial distributions of pumping, diversions and land-use management impacts in the basin. Second, such participatory processes allow for a better understanding of the drivers and constraints of each stakeholder, of the agencies and institutions being represented, i.e. what limits exist on each stakeholder's range of action. In this way, stakeholders can gain insights into the bases for their divergent viewpoints, and through increased understanding, be able to identify potential strategies to negotiate trade-offs between opposing groups.

4.1 The conceptual model: A common understanding of how the system works

One of the essential steps sometimes underestimated in the collaborative development of a model is the description and agreement on a common conceptual model of the system (Gupta et al 2011). A conceptual model of a system is the understanding of how it works and how the different components of the system interact with each other. Individuals – and especially those of us who are scientists and academicians –may often think we understand the overall system enough to develop a software model ourselves. However, our views and understanding of the system, as those of any stakeholder or individual involved in the process, are likely to be incomplete and conditioned by our background and our limited individual experience. In a collaborative and participatory process, with representation from all relevant stakeholders, all of these partial conceptual models will be shared and put in common as pieces of a collective conceptual model. Through these interactions, individual stakeholders will go through a process that has been termed *social learning* by improving their own understanding of the socio-ecological system. As the collective conceptual model becomes the basis on which decisions will be made, *sustainability learning* is the process by which actors gain shared understanding of what decisions are likely to be sustainable and which ones are not (Pahl-Wostl et al., 2007).

For the collaborative planning process to succeed it is important that everyone's partial views and understanding of the system contribute to the overall conceptual model of the system. There are currently no formalized approaches to ensure this is done properly as an initial stage. Physical scientists and modelers have often overlooked or failed to acknowledge that an effective facilitation of such stakeholder process can be challenging and falls within the domain of the human sciences practitioner. Drawing from applied anthropology, we propose a method that helps structure stakeholder participation for shaping a collective, agreed-upon conceptual model.

The Participatory Rapid Assessment (PRA, Chambers, 1994) process provides an environment wherein facilitators can pose questions or raise issues and allow stakeholders to appropriate and discuss them, expressing themselves in ways they feel more comfortable with. The efficacy of PRA can benefit from the use of tools (maps, diagrams, timelines) that help to focus discussions in which participants can contribute their information, perspectives and understanding of the reality. For example, participatory mapping, where participants can publicly draw upon their understanding of land use and water use practices, as well as the spatial linkages of water allocation in a basin, will be very visual and address potential misunderstandings in a display of social interactions. Diagrams can be of uttermost importance to learn about feedbacks across fields of study (water availability, crop production, economy) in the basin, social relationships, and vulnerabilities. The use of timelines will help understand how changes have been taking place in different areas across the watershed. The use of these tools will help develop a common conceptual model of the physical and social-economic system of the basin in an open and collaborative way. All participants will learn and benefit from this method, as long as it is properly facilitated.

If the decision-making process is to be truly coupled – including physical and human considerations - it has to look into the impacts on populations; both on economic activities and shifts in vulnerabilities. A holistic decision support system approach should seek to provide insights on different forcings in the basin, including the effects of globalization (social, economic, and environmental impacts), local manifestations of climate change impacts, and their joint effects, what O'Brien and Leichenko (2000) termed *double exposure*. To address such issues, linkages techniques such as Venn diagrams may show which external drivers may be at play in the basin and how they impact the basin system. Diagrams can show how communities see themselves integrated within the global world, their relationship with outside influences, as well as provide insights on how to become less vulnerable to external drivers.

A collectively agreed-upon conceptual model of the physical and human system of the basin will help stakeholders and decision-makers understand what are the main issues and challenges, at the basin scale and for each stakeholder. The process of putting in common everyone's understanding of the system (i.e. conceptual model) may enlighten some cause-effect relationships, as well as make evident which ones are not well understood, making evident where the uncertainties and the unknowns are in the system. These steps are essential to formulate the questions that need an answer to move forward any decision-making processes. What do we know now and what do we need to know in order to make informed decisions? Once the key questions that need to be answered have been formulated, then considerations on what type of modeling tools and decision-support systems can be pursued.

If stakeholders and decision-makers are involved in the process of developing a collective conceptual model (or shared understanding) of how the system works and what are the

main issues and unknowns that need to be answered in order to make planning or management decisions, they will likely support and invest themselves in a planning process involving the development of computer models and decision support tools.
In addition, models developed in a participatory way provide a commonly agreed upon representation of a system and its problems (Lynam et al., 2002). They become an image of the common understanding that, although imperfect, can be changed and improved with time. The participatory analysis during model development, and its contribution to decision making, brings with it the necessary social learning that can alter and inform perceptions of local problems and their cause–effect relationships.

4.2 DSS models in the middle rio grande (NM) and upper san pedro river (AZ)

In the Middle Rio Grande and the Upper San Pedro River, both DSS were developed in collaboration with stakeholder groups within the setting of an open and participatory process to solve management problems.
Following a state-wide water planning process in New Mexico, a voluntary group composed of diverse stakeholder representatives from the Middle Rio Grande planning region, and called the Middle Rio Grande Water Assembly (MRGWA) was the entity responsible for the planning. Composed of five groups focusing on agriculture, environment, urban development, water management and special technical issues, the MRGWA started a public consultation process through monthly and quarterly meetings that finally produced five scenarios or tentative management plans for the region. These scenarios comprised different sets and combinations of 44 water management alternatives identified by the public during the initial consultation processes. The quantifiable alternatives were included in the Middle Rio Grande DSS model, which allowed a quantitative comparison of the water conservation alternatives. At the end, the five scenarios were combined to form a "preferred management plan" by the MRGWA, in close collaboration with the Middle Rio Grande Council of Governments (MRGCOG), representing the local governments that would be responsible for implementing the final plan. Besides helping planners (MRGWA) and decision-makers (MRGCOG) to compare and evaluate alternatives proposed by the public, the model was instrumental to familiarize and engage the public itself in the planning process (Passell et al., 2003).
In the case of the Upper San Pedro basin, the Upper San Pedro Partnership was created to solve the management challenge in the basin and close the gap between human demand, natural availability and environmental needs. The USPP is also an organization composed by stakeholder representatives from 21 state and federal agencies as well as other entities and user groups, functioning at a voluntary basis. It is structured in three main committees: the Partnership Advisory Committee (PAC), the Executive Committee (EC) and the Technical Committee (TC). The PAC is the decision making body representing all entities; the EC represents the member entities that finance projects and operations; and the TC coordinates technical and scientific advice and oversight. Composed by representatives with technical and scientific profiles from the member entities of the USPP and the modelers from the University of Arizona, the TC reports to the PAC, so that decision-making can be science-based. The DSS model was developed through monthly open meetings with the Technical Committee, where other stakeholders and the public could participate. Representatives in the TC had to agree and decide on alternatives and conservation measures to be included in the model, as well as underlying assumptions, how to deal with

uncertain parameters and how model results should be displayed and visualized. At every meeting, the modeler would present the inclusion of last meeting's decisions into the DSS model, review them with the group and discuss the next steps of model construction, making it a collaborative, participatory and transparent endeavor (Serrat-Capdevila et al., 2009).

The Cooperative Modeling Group in the Upper Rio Grande is the equivalent to the Technical Committee in the Upper San Pedro. In both settings, these technical groups were in charge of developing and synthesizing the technical and scientific information that would be the basis of the planning process, working with the DSS model development, and other related tasks. In both cases, there was an effort to build public confidence and trust in the planning model (it properly addressed the issues at hand) as well as a sense of ownership (the model and the management alternatives were distilled from everyone's concerns and views).

Although the planning processes in the Rio Grande and the San Pedro River are the result of different institutional drivers (Statewide planning initiative in NM vs. basin initiative to meet a federal mandate in the San Pedro), the planning is structured around parallel organizations with similar roles. Although neither the MRGWA nor the USPP have any powers to impose policies or have any decision-making status, their individual member entities may have such powers within their particular jurisdictions. The understanding that comes from having to work together within a collaborative setting is key to influencing each other's work in terms of what actions are or are not sustainable or convenient. Most importantly, these planning and decision-support processes provide the opportunity to engage both the public and the actual decision-makers well before decisions need to be made. Thus the process itself, even long before the completion of the DSS product, will likely have significant positive contributions, and the way it is conducted will have important implications. The understanding of the physical system, of what is or not convenient for the common good, and of other stakeholders' needs and concerns can facilitate the finding of tradeoff solutions among competing needs.

For the interested reader, Serrat-Capdevila et al. (2009) provides an analysis of the lessons learned and the contributions of the participatory process by which the DSS model in the San Pedro basin was developed. Cockerill et al. (2006) presents the feedbacks from the Cooperative Modeling Team in the Upper Rio Grande.

4.3 Shared vision planning

There have been many efforts from varying perspectives to establish a methodological framework for science-based collaborative planning and decision-making. Liu et al. (2008) present an excellent study of integrated modeling to support natural resource management. Their work is presented from an academic perspective and a desire to improve the credibility, legitimacy and saliency of scientific information so that decision-makers use it. They frame their work within the setting of participatory processes but focus their efforts on the contributions of an integrative modeling approach. Mahmoud et al. (2009) has a broader scope, placing integrative modeling approaches as a tool to support scenario development for decision making. They emphasize the need for stakeholder input in order for the scenario analysis to be useful to decision-making.

Perhaps the most widely used participatory planning methodology in the US has been Shared Vision Planning (SVP). The main difference with respect to Liu et al. (2008) and

Mahmoud et al. (2009) is that SVP was developed and refined by planning practitioners that needed to solve planning challenges in their professional life. Authorized by the US Congress and motivated by the 1988 drought, the method initially appeared as the Drought Preparedness Study (Werrick and Whipple, 1994) with the goal of finding better ways to manage water during drought. The report is based on the joint effort of over 100 practitioners and researchers on how to approach water management issues in many case studies across the country during drought. The report highlights that drought responses are primarily behavioral and "*their success depends on people understanding their role, and knowing how their actions fit in a larger response*". It also states that planning will be much more effective if it benefits from collaboration between government agencies and stakeholders. This will provide easy access to insights and knowledge from the stakeholders (integrative plans), they will learn about the broader picture (social learning, understanding), thus being less vulnerable themselves, and will ensure public support for any potential water management plans (credibility and trust). The Drought Preparedness Study presented a methodology to set up a functional and integrated multi-stakeholder process to find planning solutions in the face of droughts, but can be used in any water management issues. The full report is available online at: http://www.iwr.usace.army.mil/docs/iwrreports /94nds8.pdf

Since its initial development, the method has been adopted by the US Army Corps of Engineers in many conflict resolution efforts in US water management regional disputes, and is commonly known now as Shared Vision Planning (SVP). SVP is based on three principles: (1) traditional and time tested planning methods and techniques (such as described in chapter 2); (2) structured public participation; and (3) use of computer models collaboratively developed in order to support the participatory planning process (Cardwell et al., 2009).

To efficiently benefit from stakeholder participation, SVP uses Circles of Influence as a way to structure involvement and engage stakeholders depending on their role in the process. As shown in Figure 1, participants can fall in Circles A, B, C or D, ideally representing the following:

Circle A: Planners and model developers. Their task is to integrate the work of others to develop planning alternatives and modeling tools to help decision-making. They form the core planning team that facilitates communication across the different circles.

Circle B: Stakeholder representatives and technical experts. Sometimes organized around working groups on specific issues, they provide information, insights and advice. They validate the work of Circle A and can evaluate proposed plans.

Circle C: The general public, whose members should have representatives in Circle B. A mechanism should exist to inform them and allow their feedback regarding the work of Circles A and B.

Circle D: The decision makers. Those who will ultimately decide what decisions are taken and what plans are implemented. They should be identified and actively engaged along the planning process, so they can provide feedback and guidance to the process.

These circles of influence are relatively natural, and they can be well illustrated by the case studies in the Rio Grande and the San Pedro basin, with slight differences. The Cooperative Modeling Team in the Middle Rio Grande and the Technical Committee of the Upper San Pedro Partnership would compose Circle A, the hands-on planners, in each basin. The Middle Rio Grande Water Assembly and the Upper San Pedro Partnership as stakeholder

consortiums as a whole would compose Circle B, providing information to Circle A and validating its progress. Circle C is the general public in both cases. Finally, the Middle Rio Grande Council of Governments and the Partnership Advisory Committee would compose the cores of Circle D in each basin, with the possibility of other decision-making agents existing beyond those groups.

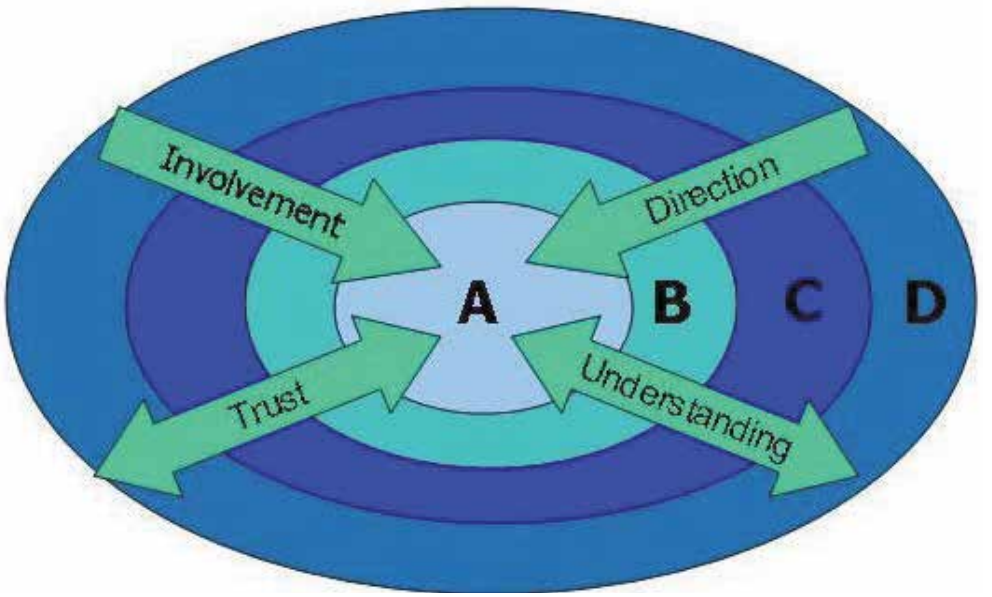


Fig. 1. The concept of the Circles of Influence (from Cardwell et al., 2009).

Currently, Shared Vision Planning is being applied in Peru, in the setting of a nation-wide water reform, prompted by the adoption of a new law in March of 2009. This law promotes the implementation of Integrated Water Resources Management (IWRM) at the basin scale through stakeholder participation, multi-sectoral integration and decentralization of planning and management to the basin level. In order to build institutional capacity for the development of IWRM plans in Peru, the World Bank and The Inter-American Development Bank are funding a pilot project – the Water Resources Management Modernization Project of Peru (PMGRH) – targeting capacity building in six pilot basins. The pilot basins are located in the Pacific coast, ridden with water management problems due to increasing economic development, population growth, and very limited water resources. In collaboration with the National Water Authority of Peru, the International Center for Integrated Water Resources Management (ICIWaRM), a Category II UNESCO Center in the US, is supporting the project through technical advice and conducting training workshops with water managers and stakeholders about participatory planning processes based on Shared Vision Planning principles.

Table 1 summarizes the seven steps of Shared Vision Planning and compares the process with the proposed frameworks of Liu et al. (2008) and Mahmoud et al. (2009) to show the similarities despite the different perspectives from academics and practitioners.

Integrated Modeling Approach *Liu et al. (2008)*	**Scenario Analysis** *Mahmoud et a.l (2009)*	**Shared Vision Planning** *Werrick and Whipple (1994)*
--	--	(1) Build a team: identify circles of influence: planners, stakeholder representatives, agency leads, advocacy groups & decision-makers. Identify Problems and Opportunities
(1) Identify and formulate the important focus questions, using science and stakeholder input	--	(2) Develop Objectives and Metrics for Evaluation
(2) Define scenarios based on focus questions and based on key external forcings, important and highly uncertain	Scenario Definition	(3) Describe current status quo: what happens if we do nothing?
		(4) Formulate Alternatives to the status quo, through broad participation
(3) Develop conceptual basis for numerical models to be built and generate data for scenarios	Scenario Construction	(5) Evaluate alternatives and develop study team recommendations (Compare alternatives against the status quo, and evaluate them with the metrics and indicators previously developed.)
(4) Develop modeling system, calibration and validation. Adjust conceptual model		
(5) Construct scenarios by deriving model inputs and collecting outputs from model runs	Scenario Analysis	
(6) Perform indicator analysis on scenario outputs: sensitivity analysis to understand main controls and uncertainty sources. Compare scenarios	Scenario Assessment	
--	--	(6) Institutionalize Plan (ensure recommendations will be acted upon, requires written agreement to act according to the findings regardless of political and administrative leaderships)
(7)Informed Decision-making	Risk Management	(7) Implement and update the Plan (~adaptive management
(8) Monitoring & post-audit	Monitoring	
(1) Repeat, new cycle	Repeat	

Table 1. Comparison of the approaches from Liu et al. (2008), Mahmoud et al. (2009) and Werrick and Whipple (1994).

5. An evolving system: Uncertainty, DSS and adaptive management

Living in a changing world, it is evident that even if planning and management are implemented as particular actions, they are an ongoing process over the long-term. Consequently, Integrated Water Resources Management is portrayed as a spiral where the implementation of past plans is monitored and the process is re-evaluated and re-directed based upon our most current, new information. In other words, we have to plan for an uncertain future, then deal with it when it becomes the present, and learn from it when it becomes the past. Such an acknowledgement is the basis of adaptive management.

Everyone knows the future is uncertain, but how do Decision Support System Models deal with uncertainty? To what extent and how is uncertainty incorporated into DSS and how is it communicated? The truth is that uncertainty is a difficult concept to work with and is often not well represented in models and decision support tools. Many systems dynamics models state as a disclaimer that the specific values provided by the model are to be interpreted as a relative measure in comparison to other alternatives, but never as absolute numbers. This is well accepted because it still allows the comparison of different management alternatives and an overall view of their impacts in the entire system. While uncertainty can be accounted for in specific model components (physical land surface and hydrologic models) once the intention to do so is there, it may be harder to represent it accurately in systems dynamics models, perhaps due to the inability to accurately represent and blend uncertainties from many different model components of the system (i.e. behavioral and socio-economic components).

There are many sources of uncertainty in simulations: uncertainty contained in the input data (climate change projections), in the model structure formulation (recharge, runoff and evaporation transformations), and arising from issues related to boundaries and scales (e.g., regionalizing soil parameters).

Uncertainty inherent to structural representations of the physical world reflects the lack of proper understanding of physical processes or our inability to represent them properly, much less crossing boundaries of scale. As an example, in basins in Arizona that constitute some of the most instrumented and studied watersheds in the world, the quantification and the spatio-temporal characterization of natural recharge into the regional aquifer remains a formidable challenge. The estimates currently used in hydrologic models are based on empirical relationships aggregated at the basin scale that were developed 20 years ago (Anderson, 1992).

When developing a DSS model, different sources of uncertainty can be represented in different ways. During a collaborative process, stakeholders and decision-makers can decide on what sources and measures of uncertainty need to be explicitly represented in the model and which ones may better be addressed through other means. For example, climate change projections are very uncertain but a multi-model envelope of uncertainty can easily be represented using the wettest and driest models (or hottest and coldest) as the extreme cases, and assuming that future rainfall (or temperature) will fall somewhere in between these extreme cases. All the projections of climate models falling within the wettest and driest models can be averaged, providing what can be used as the highest-likelihood possibility (Hagedorn et al. 2005). Such envelopes of uncertainty in inputs that drive land-surface and hydrologic models can easily be propagated or transmitted from the input variables to the output variables (Serrat-Capdevila et al. 2007). On the other hand, there are uncertainties regarding issues that are difficult to quantify but still have important impacts

on decision-making, such as changes in economic drivers, land use cover, institutions and policies. These uncertainties may be better handled through scenario development, where alternative futures – independent of our decision-making process – can be accounted for. On the other hand, information gaps identified during model development can help identify areas of uncertainty and consequently direct research and monitoring activities.

In some cases, uncertainty can be constrained and minimized to a certain extent with studies and research, but it will always be there, especially when trying to assess the future. Acknowledging uncertainty, the concept and practice of **adaptive management** presents a framework for natural resource management under uncertainty that aims at reducing uncertainty through observation during and after management interventions. In other words, adaptive management is a decision-making process that attempts to manage systems in order to maximize both the short-term benefits of management; and the gaining of new understanding to improve management over the longer term. To accomplish the second goal – learning about the system – adaptive management relies on a few basic steps:

a. Characterizing the sources of uncertainty in the system. What are the poorly understood processes in our system and where does the uncertainty arise from?
b. System observation and monitoring of system response to management actions, during their implementation and afterwards. Is the system responding to management interventions as it was expected?
c. If the system is not responding as was expected, different potential explanations can be developed and tested in future management implementations. Such explanations of why the system behaved as it did can either be consistent with our previous understanding of the system, but can also question it. Information and data gathered in future management interventions could be used to validate or invalidate such explanations. This is also known as testing assumptions and hypothesis.
d. Including and assimilating new data and information in a conceptual and numerical representation of the system, embodying the current understanding of how it functions.
e. Management can be specifically geared towards tackling domains of the system where less is known about its functioning or where major uncertainties lie. This can conflict with management goals to maximize beneficial use of the resource in the short term, but is considered a benefit for the long-term as it is likely to reduce uncertainties on the system.

Flexibility is an important aspect of a good adaptive management practice. Institutions should be able to change past policies based on the observed impacts such policies had on the system. The key to this essential feedback linking the latest observations with the next decision-making steps is that it requires close collaboration between those who monitor, study and interpret the behavior of the system with those who do the decision-making. Traditionally, these groups of people belong to different institutions, the communication among which is not necessarily fluid. It is for this reason that a true adaptive management mechanism must also foster new organisms and institutional strategies that will be able to put new knowledge to use at a practical level. For management to be adaptive, the policies must be flexible, not just the institutions.

As real-world systems are often very complex, adaptive management must make use of modeling tools to properly simulate and understand how the system functions. Ideally, as previously mentioned, this forces decision-makers, scientists and model developers to work collaboratively in a cycle of management decisions, implementation, monitoring,

interpretation of new data, and inclusion in conceptual and numerical models of the system to help validate past interpretations and/or provide new working hypothesis of how the system behaves.

To the present date, DSS models have mostly been viewed as a product that can be developed to help answer management and planning questions at a given time. It is only very recently that DSS models are starting to be perceived as evolving tools. Rather than developing and using them once, they offer greater benefits when they are dynamically changed over time to represent the evolving present, becoming a working tool that may never be a *finished product* but a product to work along the years. In participatory planning processes this allows the model to be a common representation of the system and the DSS model and supporting documentation can be like an "accountability trail" of what has been done in the past. In adaptive management practice, a DSS model will have to be updated as ongoing policies and management actions are implemented. Model updates will reflect modifications in the *engineered system layer* (canals, pipes, wells, dams, water re-allocations, changes in use efficiencies, changes in land use cover, etc.) as well as new or modified understanding gained through adaptive management on how the system works.

The issues of model updates and institutional flexibility can be well illustrated by the worries of many stakeholders in the San Pedro Basin, collected in a study to evaluate the contributions of the collaborative process in the basin. Being able to feed current, accurate and updated data into the model was a concern for the future that relates well with institutional limitations. A modeling team from the University of Arizona had ensured model and data accuracy, along with technical people from different government and state agencies involved in the process. The point was raised that if the modeling team left the collaboration, no human capabilities existed within the basins' managing institutions to easily take over and continue the modeling work. Local capacity building to update and modify the model was necessary: *Otherwise, if [the main modeler] leaves the State and stops working on it, nobody is able now to take care of things and move on from here.* A comment by one top level policy person illustrates the precarious institutional integration and the need for new flexible institutional arrangements: *"The model will help us a lot in our planning and zoning, our municipalities and county entities, water districts, water planning, etc. [...] my concern is how to keep it up to date with future science, options, and alternatives. If federal funding fails to help [the process] …if no more money comes, all will be lost."* (Serrat-Capdevila et al., 2008).

The final important point to make here is that an integrative modeling approach in adaptive management institutions will be essential in these types of contexts for many reasons. Decision makers usually use (or benefit from the use of) medium or coarse resolution models in system dynamics platforms (DSS models) that incorporate findings of more refined models in a simplified but still accurate manner. As new information and understanding becomes available, these DSS models are likely to be unsuited to the assimilation of such information. Instead, the more detailed physical models that support and inform system dynamics simulations, are more likely to accommodate new data properly and help improve the understanding of that particular component of the system. Once this is accomplished, the DSS model can be modified accordingly to accurately represent new findings in a simplified way. The full potential of adaptive management can only be reached when it is coupled with an integrative decision support systems modeling approach and with continued research and observation.

6. Conclusion

Decision Support Systems have transitioned from engineering tools to systems that provide frameworks for stakeholder participation to guide, inform and support decision making in a transparent and more sustainable way. The research and past experiences presented in this chapter have shown that participatory planning and management processes can greatly benefit from an integrative and holistic modeling approach. Models of different resolution and complexity that serve different purposes can be used to inform each other through feedbacks. While high-resolution Land Surface Models are necessary when there is a need to accommodate in detail the processes in the physical environment (such as the land-atmosphere partitioning of water and energy, the role of vegetation and the interactions between surface and groundwater hydrology), medium- and coarse-resolution models are typically better suited to modeling human interventions on the environment (such as land-use management, engineering infrastructure). Medium resolution models allow us to represent water allocation and re-distribution within the system and across uses, while coarse resolution models are used to properly describe socio-economic and institutional aspects of water management over the natural and engineered system, with a resolution at the scale of the sub-watershed. In addition to providing an efficient way to represent the coupled natural-human system, a major benefit of multiple resolution modeling is that information and findings can be readily transferred across models and used for model refinement. Information regarding natural processes, climate change impacts and feedbacks in the natural system can be up-scaled to higher level models, while behavioral and policy feedbacks from the socio-economic and institutional models can be used to drive lower resolution models and assess impacts on the natural system.

This integrated modeling approach can be the scientific foundation for participatory planning processes and the collaborative development of decision support tools. The combination of structured stakeholder participation and the use of integrative modeling will allow the proper identification of problems and management objectives in the basin, as well as a better shared understanding of the system functioning, and the development of future scenarios and management alternatives. Based on conflict resolution concepts, this methodology will not only lead to agreed-upon management solutions, but also to a well informed and educated stakeholder community in the basin. Sustainable learning comes with a better understanding of the system as a whole; and problem-solving, over the long term, can benefit from the human capital among individuals involved in participatory processes and the groups they represent. Past studies have pointed out the importance of human capital in society over economic welfare, as well as the mechanisms for ensuring it (education, research, health care, social investments), as the key quality required to address environmental and sustainability challenges. The reinvestment of resources towards human capital (knowledge) in a higher priority over economic capital can be in itself a definition of a sustainable system.

This resonates well with the learning goal of adaptive management. In the present time of rapid economic and environmental change, the future now seems to be more uncertain than ever. With the influence of climate change, the premise of a stationary state on which much of water resources planning and management are based, is now compromised. It is likely that we will have to change the ways in which we extract and use information from the past to predict the future. The implementation of efficient adaptive management mechanisms combined with integrative multi-resolution modeling capabilities will have

to balance the search for new understanding and the short-term economic benefits of management.

Currently, the main challenge to achieving efficient adaptive management remains to provide within existing institutional arrangements, sufficient flexibility and the capacity to close the feedback loop between system monitoring, modeling and scientific analysis, stakeholder participation and iterative decision-making. As this is accomplished, it will enable water resources management to shine through the lenses of economic efficiency, social equity and environmental sustainability.

7. References

Anderson, T.W., Freethey, Geoffrey W. & Tucci, P. (1992). Geohydrology and Water Resources of Alluvial Basins in South Central Arizona and Parts of Adjacent States, U.S. Geological Survey Professional Paper 1406-B, United States Government Printing Office, Washington.

Cardwell, H, Langsdale, S. & Stephenson, K. (2009) The Shared Vision Planning Primer: How to incorporate computer aided dispute resolution in water resources planning. Available online at:
http://www.sharedvisionplanning.us/resReference.cfm

Chambers, R. (1994). "Participatory rural appraisal (PRA): Challenges, potentials and paradigm" World Development, Elsevier, vol. 22(10), pages 1437-1454, October.

Forrester, J.W. (1968) *Principles of Systems*, 2nd ed. Pegasus Communications.

Gupta HV, Brookshire, DS., Tidwell, V. & Boyle, D. (2011 in press), *Modeling: A Basis for Linking Policy to Adaptive Water Management*, Chapter 2 in Brookshire D, HV Gupta, and P. Matthews (Editors), *Water Policy in New Mexico: Addressing the Challenge of an Uncertain Future*, RFF Press, Resources for the Future Book Series: Issues in Water Resources Policy Series.

Hagedorn, R., Doblas-Reyes, F.J. & Palmer, T.N., (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. Tellus 57A, 219–233.

Kang, D.S. & Lansey, K. (2011) Development of a Water Management DSS for the Upper San Pedro River Basin (To be submitted to journal).

Liu, Y., Gupta, H., Springer, E. & Wagener, T., 2008. Linking science with environmental decision making: experiences from an integrated modeling approach to supporting sustainable water resources management. Environmental Modelling & Software 23 (7), 846-858.

Loucks, D.P. & da Costa, J.R. (1991) Decision Support Systems, Water Resources Planning, NATO ASI Series G: Ecological Sciences, Vol. 26. ISBN 3-540-53097-5 Springer-Verlag.

Loucks, D.P., Stedinger, J.R. & Haith, D.A. (1981) Water Resource Systems Planning and Analysis, ISBN 0-13-945923-5 Prentice-Hall, Inc.

Maass, A., Hufschmidt, M.M., Dorfman, R., Thomas, H.A., Marglin, S.A. & Maskew Fair, G. (1962) Design of Water-Resource Systems, New Techniques for Relating Economic Objectives, Engineering Analysis, and Government Planning. Harvard University Press, Cambridge Massachusetts.

Mahmoud, M., Liu, Y., Hartmann,H., Stewart,S., Wagener, T., Semmens,D., Stewart, R., Gupta, H., Dominguez, D. & Dominguez, F. (2009) A formal framework for

scenario development in support of environmental decision-making. Environmental Modelling & Software 24 (7), 798-808.

Meadows, D.H., Meadows, D.L., Randers, J. & Behrens III, W.W., (1972) The Limits to Growth. New York: Universe Books. ISBN 0-87663-165-0

O'Brien, K.L. & Leichenko, R.M., (2000). Double exposure: assessing the impacts of climate change within the context of economic globalization. Global Environmental Change 10 (3), 221-232.

Passell, H.D., Tidwell, V., Conrad, S.H., Thomas, R.P. & Roach, D. (2003) Cooperative Water Resources Modeling in the Middle Rio Grande Basin, Sandia National Laboratories Internal Report, SAND2003-3653.

Richter, H. E. (2006). Participatory learning on the San Pedro: designing the crystal ball together. Southwest Hydrology 5(4): 24–25

Roach, J. & Tidwell, V. (2009) A Compartmental-Spatial System Dynamics Approach to Ground Water Modeling, Ground Water 47 (5), pp. 686-698, doi: 10.1111/j.1745-6584.2009.00580.x

Serrat-Capdevila, A., Browning-Aiken, A., Lansey, K., Finan, T. & Valdés, J.B. (2008) Increasing Socio-Ecological Resilience by placing science at the decision table: The role of the San Pedro Basin Decision Support System Model (Arizona). Ecology and Society 14(1): 37. http://www.ecologyandsociety.org/vol14/iss1/art37/

Serrat-Capdevila, A., J.B. Valdés, J. González Pérez, K. Baird, L. J. Mata & Maddock III, T. (2007) Modeling climate change impacts - and uncertainty - on the hydrology of a riparian system: the San Pedro Basin (Arizona/Sonora), *Journal of Hydrology* 347, 48-66. DOI 10.1016/j.jhydrol.2007.08.028

Tidwell, V. C., H. D. Passell, S. H. Conrad, & Thomas, R. P.. (2004). System dynamics modeling for community-based water planning: application to the Middle Rio Grande. Aquatic Sciences 66:1–16. DOI 10.1007/s00027-004-0722-9.

Yalcin Sumer, D. & Lansey, K. (2004) Evaluation of Conservation Measures in the Upper San Pedro Basin. In: *Critical Transitions in Water and Environmental Resources Management*, Pcdgs. World Water and Environmental Resources Congress 2004.

Wagener, T., Liu, Y., Gupta, H.V., Springer, E. & Brookshire, D., (2005). Multi-resolution integrated assessment modeling for water resources management in arid and semi-arid regions. In: Wagener, T., Franks, S., Bøgh, E., Gupta, H.V., Bastidas, L., Nobre, C., de Oliveira, Galvao, C. (Eds.), *Regional Hydrologic Impacts of Climate Change Impact Assessment and Decision Making*, pp. 265e272. IAHS Redbook Publ. No. 295.

Water Resources Council. (1983). Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies (Principles and Guidelines).

Werick, W. & Palmer, R.N. When Should Shared Vision Planning Be Used? Institute for Water Resources. Available at:
http://www.svp.iwr.usace.army.mil/resReference.cfm.

Werick, W.J. & Whipple, W. Jr. (1994). Managing Water for Drought. National Study of Water Management During Drought. IWR Report 94-NDS-8. U.S. Army Corps of Engineers. Available online at:
http://www.iwr.usace.army.mil/docs/iwrreports/94nds8.pdf

22

Decision Support System for Biebrza National Park

Ignacy Kardel[1], Jaroslaw Chormanski[1], Dorota Miroslaw-Swiatek[1], Tomasz Okruszko[1], Mateusz Grygoruk[1,2] and Martin J. Wassen[3]
[1]Warsaw University of Life Sciences -SGGW, Department of Hydraulic Engineering, Warszawa,
[2]Biebrza National Park, Goniadz,
[3]Utrecht University, Copernicus Institute, Environmental Sciences, Faculty of Geosciences, Utrecht
[1,2]Poland
[3]The Netherlands

1. Introduction

A Decision Support System (DSS) was developed in order to quantify environmental changes that occurred during the last two centuries leading to degradation of the wetland ecosystems of the Biebrza National Park. The main idea was bringing together scientific knowledge and practical experience under one framework and subsequently creating and facilitating communication between a network of scientists, nature managers and local stakeholders. As a result, an analysis tool was created, which is used for the description of relations between land use, management activities, surface- and groundwater status, conservation and restoration measures and the current state of wetland ecosystems. The system was built in 2001- 2004 in the framework of a PIN MATRA project by seven Dutch and Polish institutions, then improved in the period 2007-2010 within the cooperation of six Polish and Norwegian institutions with an EEA Grant. The previous stages of the DSS development have been described in several ways (Chormanski & Wassen, 2005; Kardel et al, 2009; Chormanski et al, 2009a).

2. Area

The Biebrza Wetlands are one of the last undisturbed lowland river systems in Europe. The Biebrza River (N.E. Poland) and its riparian areas form a large, fairly pristine inland freshwater wetland. It is a hotspot for biodiversity, which is highly valued for flora, avifauna and mammals (beaver, wolf, otter, elk). Founded in 1992 the Biebrza National Park (BNP) manages a 690 km^2 area, consisting of 160 thousand parcels (lots; half of them are privately owned). BNP is the largest of the Polish National Parks with the following land cover structure: 43% wetlands, 31% agricultural land (extensive grasslands and pastures) and 26% forests. Several publications have been devoted to the BNP description, the latest data can be found in Wassen et al. (2006).

3. Functioning of the system

The main idea for the development of the DSS for BNP was to centralize and automate the organization, collection and processing of monitoring data in a comprehensive way. The DSS is able to store data collected by permanent and periodical monitoring as well as results of model computations. Thus it improves the analytical process performed by the BNP staff or external users. For achieving these aims two assumptions were made: first, that an input data to the system would be delivered from the different source (e.g. researchers, farmers, tourists, students), and second, that input data would be stored exclusively in one place using Internet. As it is shown on Figure 1, the input to the system consist of:

- data collected by permanent and periodical monitoring conducted by BNP and other institutions;
- reports, questionnaires and random observations of whole groups of system users;
- results of calculation of different models – hydrological, ecological and others which use data from DSS.

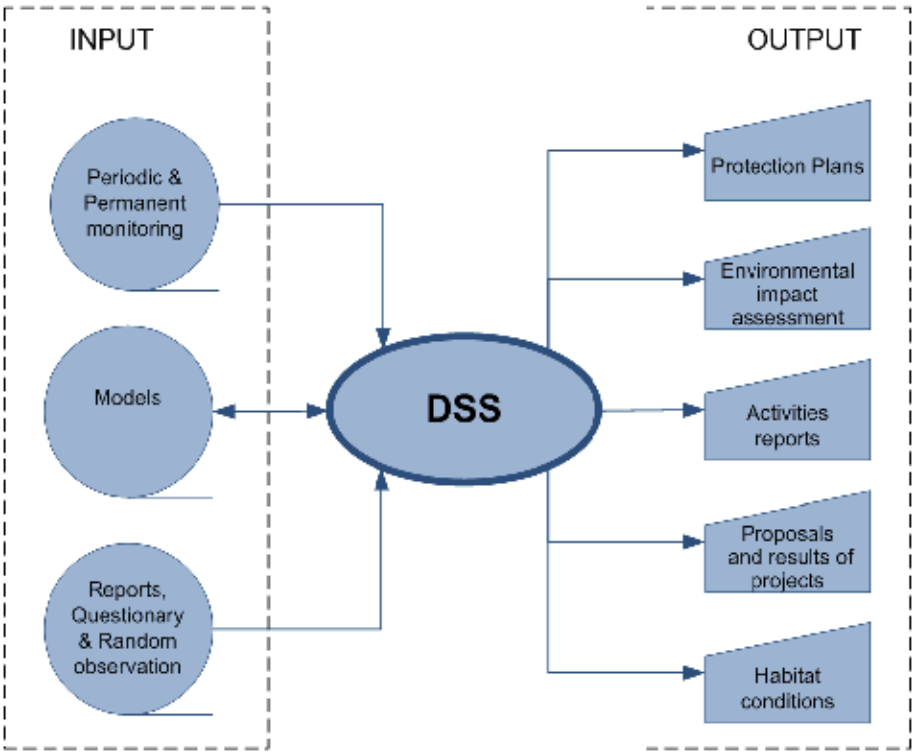


Fig. 1. General system diagram

The main tasks of DSS is verification, formatting, storing, distribution and analyzing of data considering the needs of the specific user. The user needs can be various and the system should provide answers to users in an automatic or manual way depending on a queries status - standard or non-standard. The standard queries are periodic activity reports like: list of protected species or range of protection activities. An automatic process of other nonstandard queries would be conducted depending on the needs of a user. However, it

should be pointed out that most nonstandard queries will be realized manually by using a tool such as ArcGIS Desktop. Actually, the DSS was already used for realizing several queries described in Figure 1 (frame output).

4. Users and beneficiaries

Users of the system are divided into five groups: administrators, staff, scientists, tourists and farmers, which own or rent parcels in the BNP. Additionally, staff members - about 100 people - are divided into eight subgroups (management, accounting, administration, rangers, research, education, library, extension service for farms). Activities of each group emerged from their duties imposed on the employment relationship (staff, administrators), profits (farmers) or interest and knowledge (scientists and tourists). The main activities are listed in the Table 1.

Administrator	**Park staff**	**Scientists & students**	**Farmers**	**Tourist**
Coordinating declarations of property rights on data	Iissue permission on research and land renting	Obtaining the permission for research	Browsing which land can be rented	Maps and photos review
Assignment permission	Input data by Ms Access forms	Download or connect to DSS data	Purchasing the permission on land renting	Excursion planning
Data validation, unification and conversion	Input data by mobile forms	Research planning	Crop and plant quality	Downloading of trails and points of interests
External model & monitoring data collection	Planning	Measurement	Time for field entering	
Backup system	Proposal preparation	Analysis and modeling	Terms of mowing	
Update meta-data	Project realization	Publication	What happens if you stop the mowing?	
	Periodic reports	Send modeling and monitoring results to DSS	Schedule training, meetings, rental of equipment.	

Table 1. Some of DSS users activity

A central unit of the system consists of two computer servers: web server and database server with configuration of Microsoft Windows Server, MS SQL Server, and ArcGIS Server (Fig 2). There are two main applications working parallel on the web server: the first is a geoportal which is working on Apache Tomcad, the second is ArcGIS web map services supported by Internet Information Server.

The Biebrza National Park's Local Area Network (LAN) works in a domain managed by a Microsoft Windows Server setup on second server – called Database Server or Repository. All employees of the park have access to DSS resources in the reading mode. Writing mode

access is given only to specific users who are responsible for particular database updates. Data are entered into the system through MS Access or by ArcGIS Desktop applications. Furthermore, the data are recorded directly in the field on a PDA device with Arc Pad, ArcGIS Mobile or other software.

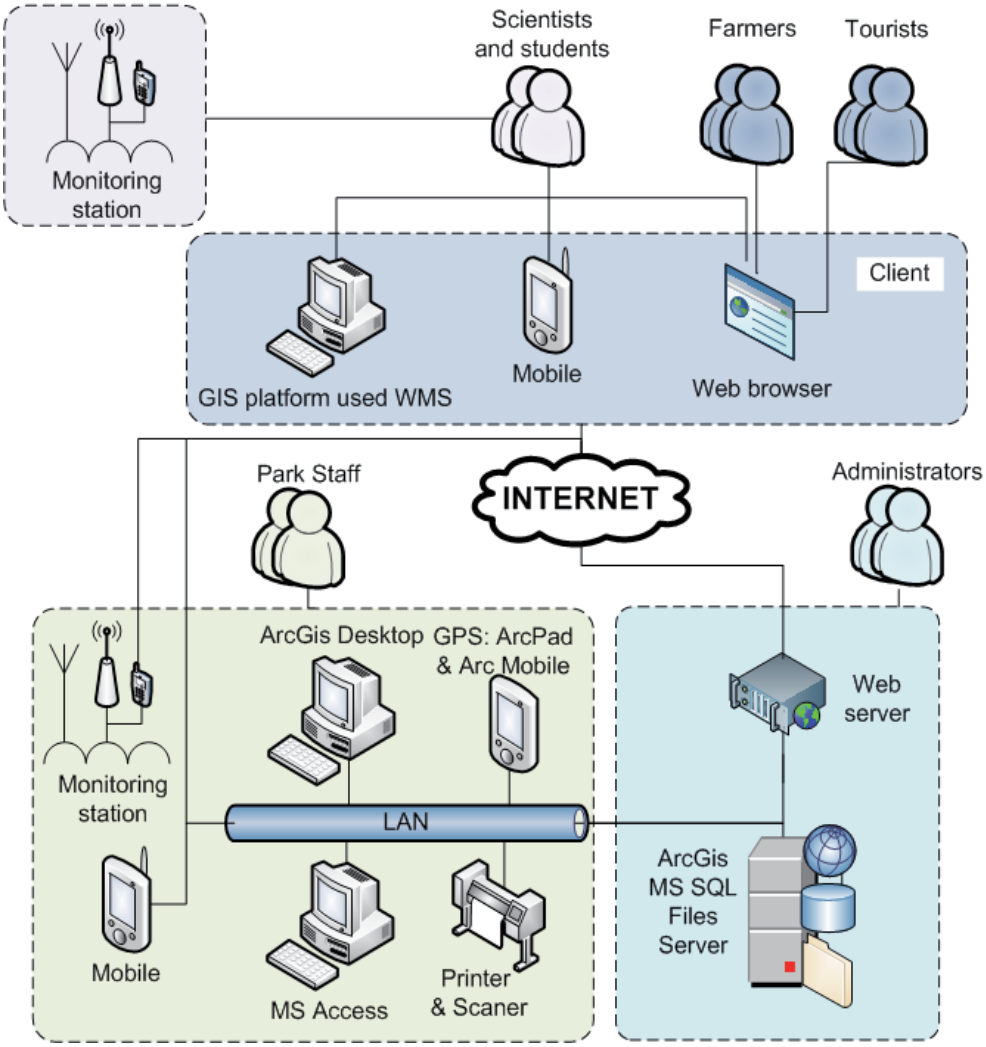


Fig. 2. Communication flow diagram for DSS

The automatic environmental monitoring system which is owned by the Park or by scientific institutions generating continuous time series of data which is stored not directly to the system. Currently, most of these stations, however measured parameters automatically without data transfer to a cellular network due to costs and avoiding the risk of loosing data when devices are stolen.

External users can communicate with the system by park staff or by the Internet. Communication over the Internet using three methods: Hypertext Transfer Protocol (HTTP), Web Map Service (WMS) and Mobile Data Service (MDS). HTTP is understood as

applications integrated on Geoportal such as a search metadata or a web map. WMS is understood as a shared map, which can be used by the client's built-in GIS software such as ArcGIS Desktop, Map Info, Quantum GIS. MDS is understood as available maps and forms prepared for mobile devices equipped with GPS receivers and wireless transfer which allows for sending measurement results via the Internet to ArcGIS Server. Moreover, the results of the automatic monitoring stations are sent using the cellular network and the Internet. This method of data communication becomes more important for the system and can be used by workers, scientists and students working with the Park. Data sent by the scientists and student is transferred to a separate database, which checks their legal status, verifies and then introduces information to the main database.

5. Monitoring

Analysis of parameters obtained from water level and discharge monitoring, provide robust and useful information on the physical stage of both water bodies and water-dependent ecosystems. Hydrological and meteorological monitoring has been placed in Biebrza National Park as an activity of special interest, indispensible for effective data-supported management. Collected and processed data of groundwater and surface water level dynamics are frequently applied in decision making processes (i.e. in issuing opinions and evaluating impact of possible investments on ecosystems of the valley of Biebrza). The data obtained from automatic and manual field measurements become also the key input to hydrological models, which broad extent in particular basins of the valley allowed to interpret the status-quo of ecosystems and to analyze certain scenarios of their development in time and space. A hydro-meteorological monitoring network of the BNP was established initially in the very first years of the Park functioning (1990ies) and then extended by different scientific institutes conducting research in the Biebrza wetlands. As hydrological processes play a siginificant role in the valley of Biebrza for wetland ecosystem development, both surface water and ground water have been continuously monitored since 1995 on numerous locations. The current monitoring network is primarily supervised by the BNP, Institute of Meteorology and Water Management and Warsaw University of Life Sciences staff. It consists of 25 surface water gauges, more than 120 groundwater piezometers, three automatic meteorological stations and four rain gauges. Continuous dataset from monitoring network can be additionally supplied with data collected within the area of BNP by other scientific institutions. The monitoring network of the BNP, Warsaw University of Life Sciences (WULS) and the Institute of Meteorology and Water Management is presented in Figure 3.

Surface water monitoring and the rain gauge standard examination have been done continuously in a 1-day time-interval, however for certain research purposes the surface water network in the lower Basin is running with 6 hours interval, while in upper Biebrza River automatic water level sensors are recording water levels with a 10 minutes interval. Groundwater level measurements are being done in approximately 10-day time interval, however in some locations automatic water level sensors have been installed in piezometers and programmed to a measurement interval of 6 hours. Automatic meteorological stations record parameters of rainfall, wind speed, air humidity, temperature and solar radiation continuously, in 1 – 30 minutes time interval.

Data of hydrological and meteorological monitoring are being collected and delivered to specialists responsible for data processing and storage by field services of the BNP and WULS researchers.

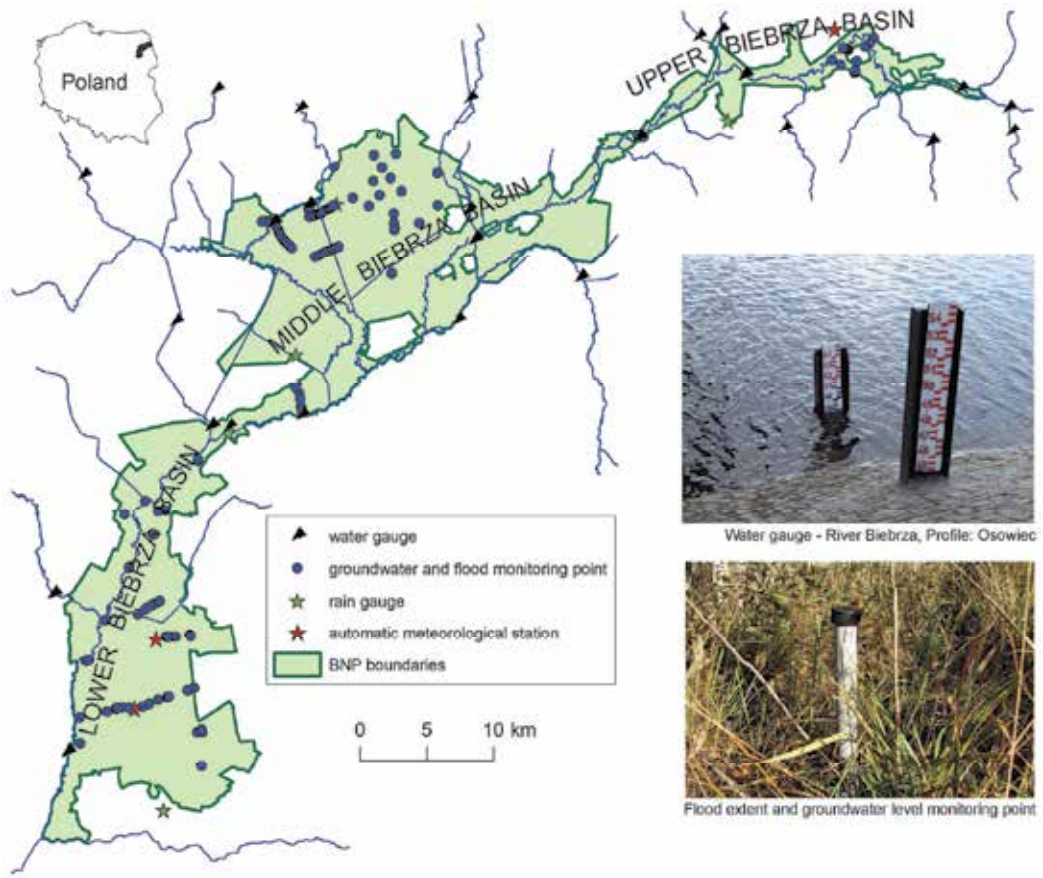


Fig. 3. The hydrological and meteorological monitoring network of the Biebrza National Park

6. Modeling

Within near-natural wetlands of broad spatial extent, most ecological dependencies and management activities and strongly linked to hydrological processes. Among ecosystems of the Biebrza Valley, groundwater discharge-dependent mires and riparian wetlands are dominant types of habitats (Okruszko, 1991). In several parts of the wetlands hydrological information was crucial for the decision on conservation or restoration activities. Therefore, hydrological models have to be set up and aimed to compute either surface-water or groundwater flow dynamics. In this regard, a number of modeling studies were applied in the Biebrza Valley (Batelaan & Kuntohadi, 2002; Grygoruk et al., 2011; Kubrak & Okruszko, 2000; Querner et al., 2010; Swiatek, 2007; Swiatek et al., 2008).

To analyze hydrological processes on groundwater discharge-dependent mires (some) models were set up on the basis of available computation algorithms (SIMGRO, MODFLOW). Hydrodynamic surface-water models (fluviogenic models) are used to analyze hydrological processes on riparian wetlands. Due to the complexity of the problems which were analyzed these models were created and dedicated only to be applied in

particular stretches of the Biebrza Valley. Results of selected modeling approaches were included in the DSS and applied in habitat contiguity and evolution analysis.

6.1 Soligenic mires models structure

To quantify groundwater discharge as a crucial factor for the functioning of groundwater fed mires, the MODFLOW model (McDonald & Harbaugh, 1988) was applied in two separate set-ups: in the Upper Basin (Batelaan & Kuntohadi, 2002; Van Loon et *al.*, 2010) and in the Middle Basin (Okruszko, 2005; Grygoruk et *al.*, 2011). Results of the modeling were analyzed due to habitat contiguity in conditions of various hydrological flow processes and fluxes in the peat soils. Both approaches pay significant attention to groundwater discharge mapping, the crucial factor for the functioning of throughflow mires and their evolution. Groundwater models compute continuous groundwater levels. Applied packages (especially the MODFLOW Drain and Seepage Packages) allowed also to interpret groundwater discharge spatial distribution patterns.

6.2 Fluviogenic wetlands models structure

The state of fluviogenic (riparian) ecosystems is dependent mostly on the conditions of their hydrological alimentation by flooding waters. The most important hydrological characteristics, conditioning of the growth and the development of swamp vegetation, are: inundation surface, mean inundation depth, the frequency and the duration of inundation. A hydrodynamic model coupled with GIS techniques makes it possible to obtain necessary data for the determination of the above-mentioned hydrological characteristics. It is also a tool which facilitates the estimation of the influence of different river valley management methods on hydraulic conditions of water flow. It can even be used as a research tool for executing effective policy of natural values protection within the Biebrza National Park.

Unsteady flow in natural rivers is usually treated as a one-dimensional flow in practice and is based on St Venant equations. In order to simulate flood flow in the lower Biebrza Basin, the one-dimensional unsteady model flow RIVER-SV was applied (Swiatek, 2007). The basic form of the non-linear St Venant equations combined with retention effects of the vegetated areas on flood wave conveyance were used in the model.

The model uses the Darcy-Weisbach relationship for the description of flow rules. It also enables introduction of water mass and momentum exchange processes between the main channel and floodplains, and it has the flexibility to account for parts of a cross section covered with vegetation and those with no vegetation. Thus, the developed model enables accounting for unsteady flow and flow resistance resulting from both vegetation covering a cross section and momentum exchange between the main channel and floodplains.

In the topological discretization scheme of the flow, the Lower Biebrza Basin and its floodplain are represented as a one – dimensional channel from Osowiec gauge to Burzyn gauge. The Wissa River (a tributary) is treated as a point lateral inflow and described by the flow hydrograph at the Czachy gauge. Geometry of the river channel and floodplain is described by 47 river cross sections. The cross sections were measured by manual sounding for the main channel part and the topography of the floodplain was calculated from the Digital Elevation Model. Field monitoring in the Lower Biebrza Basin proves that during flood periods, the river valley consists of parts which mainly act as storage for flood water and of active flow areas (Chormanski and Miroslaw-Swiatek 2006). The 1D model is capable of describing flood conditions using the appropriate geometry of cross-sections, which are limited to the active flow zones. In the developed model, the particular areas for each cross-

section were identified according to a vegetation map and topography determined with the Digital Elevation Model (DEM). The Lower Biebrza Basin model was successfully calibrated and verified for the measured data and historical data of flood events (Swiatek et al., 2008). The water level values calculated with the numerical model of flood-flow for cross-sections were used to determine the digital model of the floodwater table in the valley. Then, flood extent maps were calculated for the whole area of the valley by overlaying the DEM and water table layers.

7. IT components

Due to a historical legacy of the system, it is partly based on the relational databases (MS Access) which is converted as necessary to the Environmental Systems Research Institute (ESRI) geodatabase format. In addition, due to the low bandwidth Internet Web Server, the Park has been tested and exhibited outside its headquarters on the server at the Warsaw University of Life Sciences – SGGW (WULS). The system was tested in the BNP with particular reference to the Red Bog Reserve (a raised bog area located in the northern part of the Middle Basin) .

7.1 Databases and maps

The system consists of ten thematic databases built in a client-server structure. The database tables are located on the server while the forms are on the personal computers of individual staff. Each database is built, or is combined with a layer of information in a shp format or geodatabase by ESRI. The whole system uses software MS SQL Server, MS Access and ArcGIS Editor. The thematic databases contain information gathered from different research activities conducted since the establishment of BNP in 1993, biota and a biotic resources inventory, which was done during BNP management plan development in the late 1990ies as well as historical and current management practices recorded by the park staff.
In the system the following database are included (Chormanski & Wassen, 2005):

- Ownership database

This database contains information on parcels situated within the BPN area. Information on owners, usage and territorial units (commune, districts) is kept for each parcel.

- Non-Forest Management Activities database

This database is adapted for storing information on performed and planned protective treatments.

- Forests database

This database is adapted for storing information on forest sub-sections within the BPN area. It contains particularly information on general characteristics of more than 4400 forest sub-sections, forest valuation description, information on gaps, planned and performed protective and growing treatments, and on cuttings.

- Fauna database

This database is adapted for storing information on location, number and living environment of vertebrates and invertebrates. It contains also characteristics of species as far as relation to the human kind, zoography and ecological classification is concerned.

- Flora database

This database is adapted for storing information on occurrence and characteristics of plant species, as well as on planned and performed protective treatments.

- Water quality database

This database is adapted for storing information on location and chemical and physical properties of tested water taken from ground bore-holes, as well as from lakes and rivers. It generally contains data from the area of BPN and from its protection zone.

- Pollution database

This database is adapted for storing information related to observation of the following types of pollution: surface water pollution (direct sewage discharges), groundwater pollution (waste dumps, fertilization, dunghills), air pollution, extra-ordinary environmental hazards.

- Hydrology & Meteorology database

This database is adapted for storing information related to the following observations made at piezometers, gauge stations and weather stations: location, levels of surface water and groundwater, discharge of surface water, precipitation quantity, air temperature and other meteorological data.

- Soils database

This database is adapted for storing historical information on the following observations from soil pits, as well as cyclical surveys of physical and chemical properties of soil at various depths. It particularly contains: location, plant coverage, depth to water surface, characteristic of soil layers, general soil type, physical and chemical properties of soil.

- Fires database

This database is adapted for storing information on location of fires, causes for their occurrence, duration, number and type of fire brigades participating in extinguishing them, character and costs of damages caused by fire.

7.2 Mobile application

To improve the data collection process in DSS, recently mobile applications were developed. These applications have been successfully tested by researchers in extreme conditions of remote area, so that, after positive experience, they were distributed to the staff of the BNP. Four mobile applications have been developed: a "piezometer", "soil pit", "fauna observation" and "phytosociological photo". The task of "piezometer" application is to record water level and physico-chemical parameters of water taken at different levels, i.e.: pH, electro conductivity, temperature. However, "soil pit" application is designed to describe the habitat in the individual soil layers. It consists of a series of subforms describing in turn: place and time of observation, layer (thickness and type of soil) other physico-chemical soil parameters. Both applications were developed using ArcPad Application Builder and work in an environment of ArcPad (Fig. 4). After testing these applications, it became clear that ArcPad because of its numerous functions is complicated to use for ordinary BPN staff without specialized training. Therefore, other applications were developed in an ArcGIS Mobile (Fig. 5). "Fauna observation" application consists of one form of recording: who, where, when saw a particular animal species, in which number and which behavior. "Phytosociological photo" application consists of two forms of registration: who, where, when and as a second: in what plant community was the observation and counting performed and description of plant species. The advantage of ArcGIS Mobile application is its full integration with ArcGIS Server, which enables anywhere, anytime, to download the current state of the database of maps and forms over the Internet and in the same way to send the work on the server. This application consists of five tabs Map View, Collect Features, Search Features, View Work List, Synchronize, the corresponding status in accordance with their name on various activities.

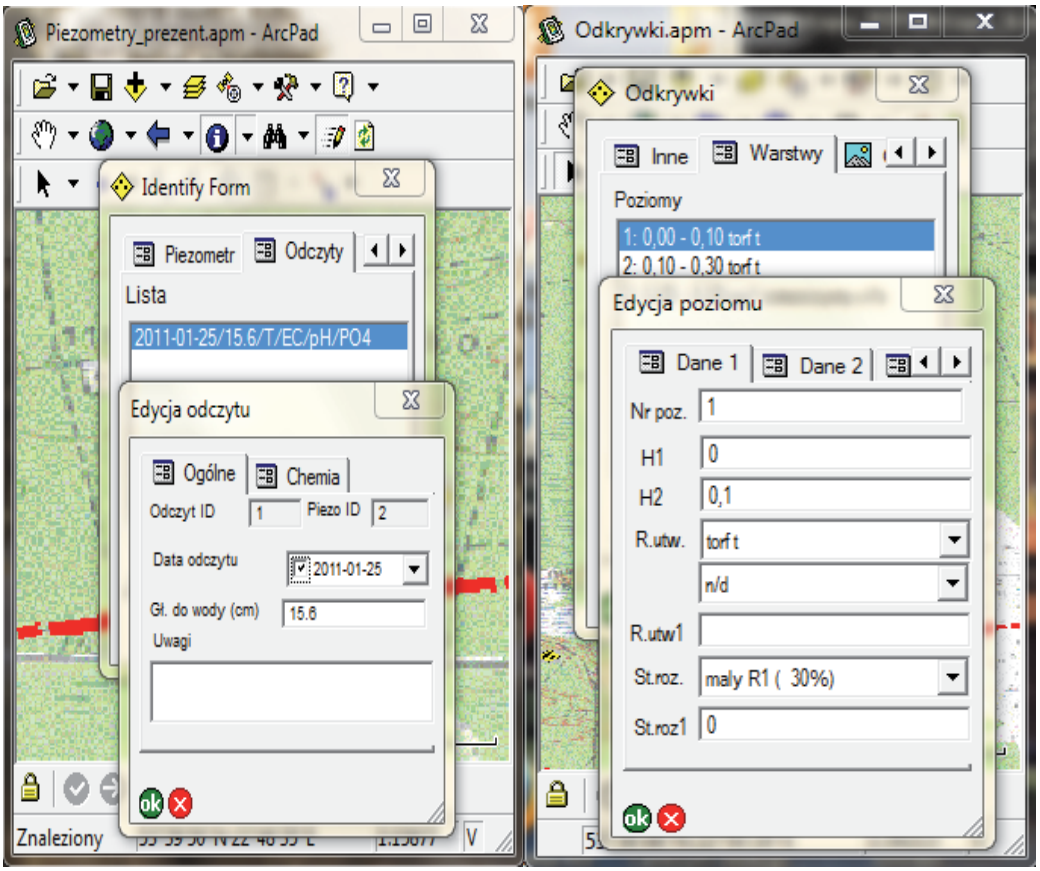


Fig. 4. “Piezometer“ (on left) and “soil pit “ (on right) forms run on ArcPad environment

Fig. 5. "Fauna observation" (on left) and "phytosociological photo" (on right) forms run on ArcGIS Mobile environment

7.3 Geoportal

From the outside user point of view the most useful tool is the geoportal, installed temporarily at http://levis-map.sggw.pl/geoportal_biebrza. This tool was created based on ESRI's product called Geoportal Extension cooperating closely with ArcGIS Server (ESRI, 2011). It is divided into five tabs: home, search (Fig. 6), browse, download, interactive map (Fig. 7) all prepared in two language versions (Polish and English). Data can be viewed in a thematic directory, located in the Browse tab. The Download tab is used to send data from selected areas by mail. However the Interactive Map tab allows to view resources, which includes information about location and basic characteristics for particular research. This tab also provides access to photographic documentation (in the format 3-D) of the studies carried out in strict nature reserves, inaccessible for tourists.

Additionally the portal has functions of registration for new users, which extends their rights for input description (metadata) of the research carried on the area of BNP. To the special features of this webpage, one can edit and search the metadata in a format compatible with the European Commission (EC) INSPIRE Directive (EC, 2007). Future plans for the geoportal is to add two new tabs: Mobile Applications and Scenario. The Mobile Applications will be used to download on PDA forms while the Scenario tab will querying the system in order to obtain answers to various scenarios.

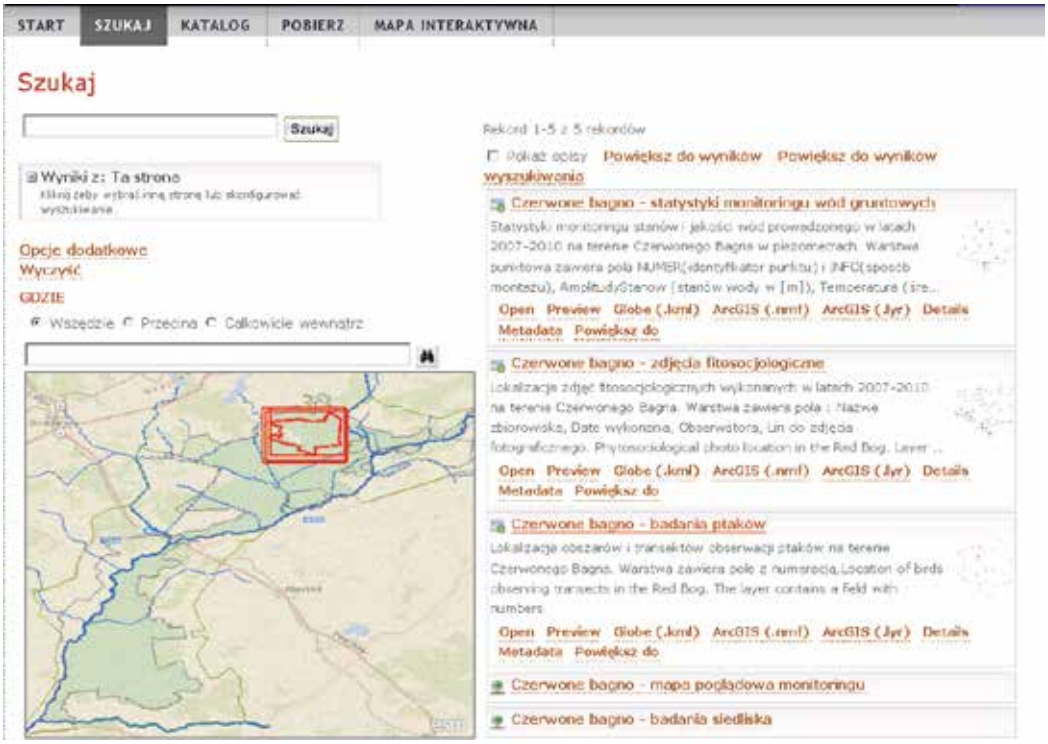


Fig. 6. Geoportal search metadata tab

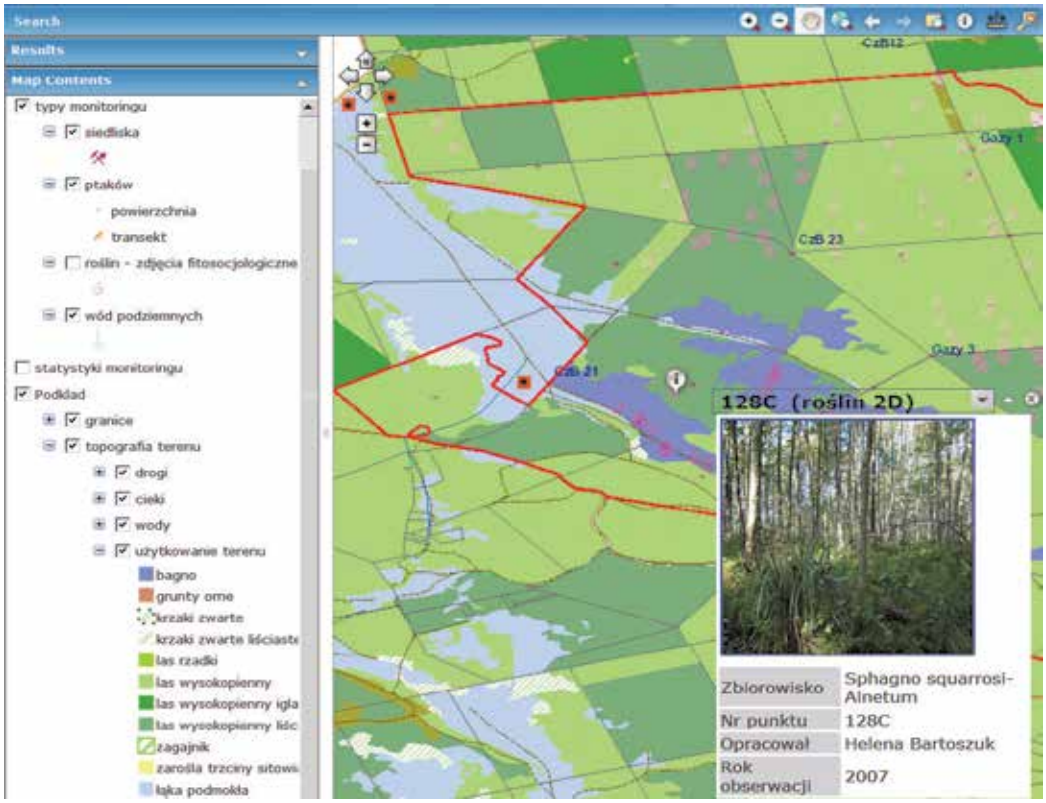


Fig. 7. Geoportal interactive maps tab

8. Stages of system development

The construction of DSS was done in a few phases: retrieving the existing data, designing the system, programming, testing and users training. The detailed scope of the work can be described as follows:

- construction of the system structure, purchase of hardware and software necessary for installation of DSS,
- creation from scratch of 10 thematic databases and five in the form of geodatabases,
- conversion of the existing descriptive and map data to databases,
- conducting field studies to identify the parameters of the models,
- simulations of several scenarios developed in Hydrological and Ecological Modules,
- training of employees appointed to use the system,
- developing, training mobile applications for data input,
- testing the applications in the field,
- development of web maps and web data access page.

9. Results

Already several studies were conducted with the assistance of the DSS. Most of them were site specific and problem oriented. The results obtained during these studies, when stored

and processed in the DSS can be of use also for other types of studies which are or will be needed for wetland ecosystem conservation or restoration, if these do not require the need to re-run models, since this is often not user friendly or requires the authors (developers) assistance. The examples given below show a range of results possibly to be obtained from different sources.

9.1 Soligenic mires models results

Results of the reference model runs as well as runs that include the input data of hydrological stress caused by human and natural impacts, are stored as raster and vector files, which make them ready to use from the level of database in GIS. Depending on actual needs and situation, new model runs can be done and analyzed as some elaborated and calibrated computation algorithms are also stored in the database. Continuity and ecological function stability of wetlands in the Middle Biebrza Basin are strongly dependent on minerotrophic groundwater discharge from the sandy aquifer to superficial layers of peat. Common balance between lateral groundwater inflow and accumulation of precipitation water on top of the peatland is responsible for the development of transitional mires. Modelling of groundwater discharge in various hydrological scenarios, that include increased groundwater inflow to the valley from adjacent plateaus (Fig. 8), can become an important verification of management strategies held by the BNP authorities. The relevance of such a modelling study increases, if a complex set of different restoration strategies needs to be evaluated. Then, some hydrological indicators that become the main force for wetland development, can be visualized and analyzed in coherence with other aspects, that particular strategy should have fulfilled.

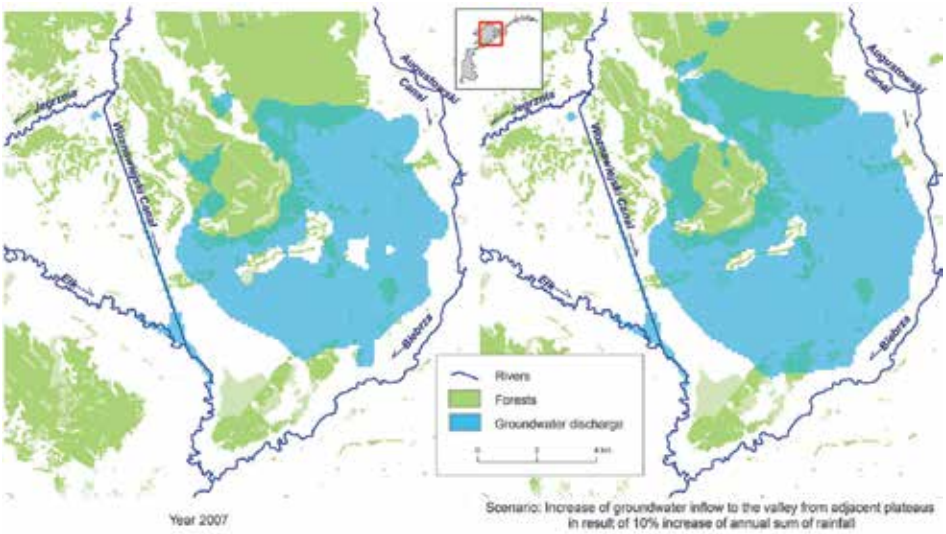


Fig. 8. Results of the steady-state groundwater discharge modelling – comparison of hydrological conditions in Middle Biebrza Basin in 2007 and in a scenario, that assumes 10% increase of groundwater inflow to the Biebrza Valley from adjacent plateaus.

9.2 Fluviogenic wetlands models results

This procedure allows for flood simulations with the hydrodynamic model and GIS analysis for determinations of inundation extent and was used for elaboration of different kind of

maps, used in the DSS system (Fig.9). This type of performance combined with the physical and chemical studies of water are the basis for the analysis of biodiversity in wetland areas.

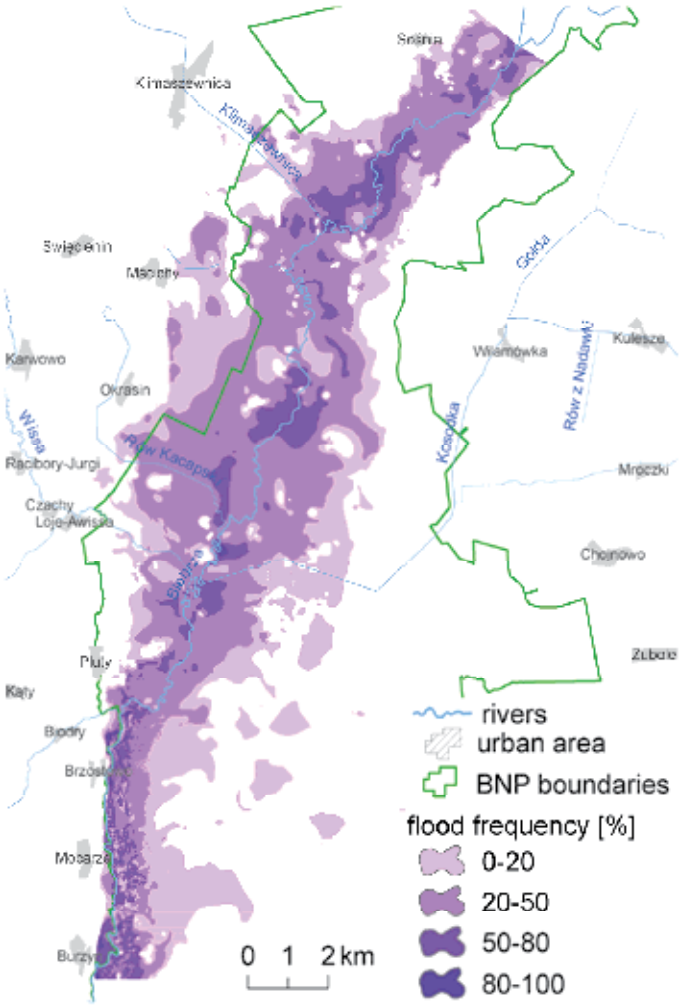


Fig. 9. Example of fluviogenic model results presenting of long-term annual mean flood frequency on Biebrza Lower Basin in the vegetation season (February-September) in the period 1961 till 1996 (Chormanski et al., 2009b)

9.3 Geostatic relationship of hydrology and ecology in riparian wetlands

The main purpose defined for the surface water flow model has been to generate suitable hydrological data for the ecological module and the data for producing thematic maps with the range of flooding for the DSS module. The long term flood characteristics are very important for ecosystem development. Chormanski et al. (2011) described significant relations between the distribution of plant communities and the extent of average floods, which could be considered as long term average inundation conditions. One of the practical

applications of the DSS is analysis of the spatial relation between flood's statistic maps and vegetation maps. The analysis is performed in ArcGIS. It shows strong relations between inundation frequency maps and different water dependence classes of ecosystems in the Biebrza Lower Basin (Chormanski et al., 2009b). The hydrodynamic module of the DSS is used as a tool for studying the relation between flood characteristics and riparian ecosystems including prediction of the ecosystem changes due to human management and climate changes.

10. Conclusion

Establishing the DSS in the BNP has increased awareness of the value of information for management purposes. This is specially important for developing of management and conservation plans as well as for preparing the application for different financial grants supporting conservation measures.

The design of the modern DSS results from the process of implementation of the EC INSPIRE Directive, under which such tools as Arc GIS Server have arisen. Implementation of new technologies has and will continue to further accelerate the process of collecting and sharing data. Moreove, the DSS will have an impact on efficient data collection and use i..e. increasing the level of accuracy of analysis and reduce their costs by optimizing the efficiency. Running the geoportal allowed the wider community to participate in the construction of the DSS and will likely improve the quality of scientific studies and projects performed in the Park. It will also encourage scientific institutions to provide the results of modeling in a digital form and will help the Park staff in decision-making in difficult to predict processes.

11. Acknowledgment

We acknowledge the support by the Dutch government for funding the PINMATRA project no 2001/039 and the support by a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and the Norwegian Financial Mechanism - Biodiversity protection of Red Bog (Czerwone Bagno) - relic of raised bogs in Central Europe - PL 0082. Research work financed by resources planned on science in years 2007-2010 as research project.

12. References

Batelaan, O. & Kuntohadi, T. (2002). Development and application of a groundwater model for the Upper Biebrza River Basin, *Annals of Warsaw Agricultural University, Land Reclamation*, No. 33

Chormanski, J. & Wassen, M. J. (Eds.), (2005). Man and nature at Biebrza; integration and dissemination of knowledge for sustainable nature management. PIN-MATRA Final Report 2001/039, In: *University of Life Sciences Warsaw / Utrecht University*, 25.02.2011, Available from: http://levis.sggw.pl/DSS/pl/endraport.pdf

Chormanski, J. & Miroslaw-Swiatek, D. (2006). Monitoring of the spring flood in the Lower Biebrza Basin, *Polish Journal of Environmental Studies*, Vol.15, No. 5D, Part I, pp. 122-126

Chormanski, J. & Kardel, I. & Swiatek, D. & Grygoruk, M. & Okruszko, T. (2009a) Management Support System for wetlands protection: Red Bog and Lower Biebrza Valley case study. Hydroinformatics in Hydrology, Hydrogeology and Water Resources, *IAHS Publ.,* No. 331, pp. 423-431, Hyderabad, India, September 2009

Chormanski, J. & Swiatek, D. & Michalowski, R. (2009b). A hydrodynamic model coupled with GIS for flood characteristics analysis in the Biebrza riparian wetland, Oceanol. Hydrobiol. Stud., Vol. 38 No.1, pp. 65-73, ISSN 1730-413X

Chormanski, J. & Okruszko, T. & Ignar, S. & Batelaan & O. & Rebel, K. & Wassen, M.J. (2011). Flood mapping with remote sensing and hydrochemistry: a new method to distinguish the origin of flood water during floods, *Ecological Engineering,* Accepted 2011

EC (2007). Directive 2007/2/EC of the European Parliament and of the council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), In: *Official Journal of the European Union,* 25.04.2007, Available from: http://eur-lex.europa.eu/ LexUriServ/LexUriServ.do?uri=OJ:L: 2007:108:0001:0014:EN:PDF

ESRI (2011). Geoportal Extension 10, *Environmental Systems Research Institute*, 25.02.2011, Available from http://resources.arcgis.com/content/geoportal-extension

Grygoruk, M. & Batelaan, O. & Okruszko, T. & Miroslaw-Swiatek, D. & Chormanski J. & Rycharski M. (2011). Groundwater modelling and hydrological system analysis of wetlands in the Middle Biebrza Basin. *Geoplanet: Earth and Planetary Sciences*, No. 4

Kardel, I. & Miroslaw-Swiatek, D. & Chormanski, J. & Okruszko, T. & Wassen, M.J. (2009). A Water Management Decision Support System for Biebrza National Park. *Environment Protection Engineering.* Vol.35, No.2, pp.173-180

Kubrak, J. & Okruszko, T. (2000). Hydraulic model of the surface water system for the Central Biebrza Basin (in:) Mioduszewski W. and M. J. Wassen (eds.), Some aspects of water management in the valley of Biebrza River, *IMUZ*, Falenty, Poland.

McDonald M. G. & Harbaugh A.W. (1988). *A modular three-dimensional finite-difference ground-water flow model*, U.S. Geological Survey Techniques of Water-Resources Investigations, book 6.

Okruszko, H. (1990). Wetlands of the Biebrza valley; their value and future management. *Polish Academy of Sciences, Warsaw, Poland*

Okruszko, T. (2005) Hydrologic criteria for wetlands protection, Dissertations and Monographs Science SGGW Publisher, p. 151, Warsaw, Poland

Querner, E. P. & Mioduszewski, W. & Povilaitis, A., & Slesicka A. (2010). Modelling Peatland Hydrology: Three Cases from Northern Europe, *Polish Journal of Environmental Studies* Vol.19, No. 1

Swiatek, D. (2007). Unsteady 1D Flow Model of Natural Rivers with Vegetated Floodplain. *Publication of the Institute of Geophysics*, No. E-7 (401), pp. 237-244.

Swiatek, D. & Szporak, S. & Chormanski, J. (2008). Hydrodynamic model of the lower Biebrza river flow - a tool for assesing the hydrologic vulnerability of a floodplain to management practice, *Ecohydrology & Hydrobiology,* Vol. 8, No. 2-4, pp.331-337, ISSN 1642-3593

Van Loon, A. H. & Schot, P. P. & Griffioen, J. & Bierkens, M. F. P. & Batelaan, O. & Wassen, M.J. (2009). Throughflow as a determining factor for habitat contiguity in a near-natural fen, *Journal of Hydrology*, Vol. 379

Wassen, M.J. & Okruszko, T. & Kardel, I. & Chormanski, J. & Swiatek, D. & Mioduszewski, W. & Bleuten, W. & Querner, E. & El Kahloun, M. & Batelaan, O. & Meire, P. (2006). Eco-Hydrological Functioning of Biebrza Wetlands: Lessons for the Conservation and Restoration of Deteriorated Wetlands, *Ecological Studies, Springer-Verlag*, Vol. 191, pp. 285-310

23

Interactive Localisation System for Urban Planning Issues

Roberto De Lotto
University of Pavia
Italy

1. Introduction

The paper discusses the benefits of specific user's oriented tools in urban planning processes, throughout the applications of an original system devised to define the localisation of urban services based on traffic distribution. Operative examples concern the Italian city of Pavia.
The system ULISSE (Urban Location Interactive System for SErvices), created by the author, uses a static model of network optimization which minimizes the global access cost (in term of time) faced by clients of a predefined service location. With a client-server structure and a friendly user interface, ULISSE provides in one single process all the classical results such as minimum paths research, definition of services influence area and access time to reach each service.
ULISSE is useful for planners to test traffic and environmental effects of localisation hypotheses, or to find the best localisation solution for specific urban functions in terms of accessibility. The paper shows examples of related applications on the city of Pavia: check of traffic pressure on environment; examination of efficiency of a new public mobility service (bus); best localisation of emergency services.
After a brief presentation of the cultural background about localisation models and urban planning, author describes: the analogical network, the calculation model, the system ULISSE and its applications.

2. Background

When approaching to regional and urban planning it is necessary to control a wide group of issues, themes, techniques and models. After the unpredictable transformations that the industrial revolution caused to the cities, modern urban planning was born as a new discipline that took its basis from architecture, engineering, hygiene, social science, aesthetics (Gabellini, 2001).
As any design process, urban planning is an extremely involved activity characterized by an iterative operative procedure strongly related to the designer's experience and know-how. For this reason, it cannot be described as a fixed series of phases, since even at intermediate steps, some corrections or modifications of the initial stage could become mandatory. Moreover, in the last decades, planning activity is oriented to divide the structural/strategic phase from the operative one (Mazza, 1997). With reference to the first phase, to define a

general shape of the territory (at the urban or regional scale), planners have to access to specific models and data processing systems with friendly user interface, while it is not essential to have very detailed and precise outputs. In other words, at the beginning of planning process it is important to have the order of magnitude of a complete variety of phenomena, which in the following phases can be completed in an exhaustive way.

2.1 Urban planning and location problems

Urban planning is the instrument to improve the best distribution of human activity on the territory (Benevolo, 1963). Planners developed localisation theory at the same time with the analysis of urban models since the mono-centre scheme deepened by Von Thünen (1826). His model of land use, that connected market processes and land use, was based on the assumption of maximizing the profits by the actors (farmers). Localisation was closely connected to the physical distance among different land uses. This deterministic and static concept took its stands on the principle of equilibrium status.

As a simplified scheme, it was related to a specific problem, and localisation itself must be considered as a particular theme into a wide-ranging and integrated model. However, at the urban scale, localisation theory always referred to the most relevant topics that strongly characterized the growth of the cities: first the agriculture land use, then the industrial areas and the big residential districts, until the commercial facilities and the distribution of urban public services (the service-oriented city).

Until 1954, general localisation theory has been related to a specific geometrical scheme (MacKinder in 1902 and Weber in 1909, which Christaller developed in his Central Place Theory in 1933), always in relation to new cities addressed to industrial settings, showing a restricted application domain. In 1954 Mitchel and Rapkin put in rigid connection land use and urban traffic also considering the modal split; this idea was used in the transportation plans of Chicago and Detroit, showing its practical use in city planning.

The Gravity Model by Lowry and Garin (developed in 1963), applied in Pittsburgh, offered a new understanding both in the field of a general urban model and in the definition of two sub-models regarding economic topics and localisation (assigning to this last theme a specific role). Even considering Von Neumann's and Morgenstern's criticism of the rational research of optimal solutions with the introducing of the concept of sub-optimal, urban models concerning location like problems spread in the last 40 years in many real contexts at urban and regional scale.

It is important not to merge localisation models with Operational Large Scale Urban Models (OLSUMs) that are mathematical simulation models describing comprehensive urban systems in great details (spatial, with a fine zoning of the territory, and functional, with a disaggregated account of urban activities and infrastructures)[1].

2.2 Localisation and accessibility in consolidated or historical cities

Application of location and accessibility models to real contexts implies the definition of the invariant conditions: consolidated and historical cities can not be warped with new mobility

[1] Born in USA late '50s of the last century, these models flourishes at best in '60s and '70s, especially in UK in a large number of applications for sub-regional territorial planning. The critics of Lee in 1973, in his very famous paper "Requiem for large-scale models", helped the lost of interest in '80s and '90s. Since the end of the XX century a slow but growing interest can be remarked.

nets (neither underground for archaeology rests nor in elevation for landscape facets) and the search of the best location for destination places (such as public services) becomes as crucial as the improvement of the existing mobility net.
However, saturation of the built territory places further problems connected to the settled urban texture, when searching new locations for different functions (i.e. impossibility to change the land destination, cost of the area, compatibility among functions and existing buildings); in this context, it is inadequate to analyze an ideal distribution of functions based on a perfect mobility net.
For these reasons it is crucial to find sub-optimal solutions considering all the existing bonds, taking into consideration the best benefits for citizens; and citizens themselves try to attain the same goal. In fact, auto-organisation of all urban systems is nowadays a milestone of urban theories (Bertuglia, 1991, Donato and Lucchi Basili, 1996). Auto-organisation in some way can be associated to the old general notion of level of satisfaction on the same indifference curve from the Bid-Rent Theory (Alonso, 1964).

3. General description

Considering the cultural background, the system ULISSE (Urban Location Interactive System for SErvices) here described is based on a model developed to help the planners in finding the sub-optimal location for public services, considering the access time (that the users spend) as a function to be minimized (De Lotto and Ferrara, 2001).
The system is based on a static model associated with the analysis of location-like problems in an urban context (Drezner, 1995). This kind of problem is usually stated as follows: given m clients and n potential sites for locating prespecified facilities, taking into account the profit deriving from supplying the demand and the cost for setting up the facility in question, select an optimal set of facility locations (Dell'Amico et al. 1997).
In order to face the problem, a classical way is to define a global cost function associated with a certain location decision to be minimized. The cost function (named C in this paper) can rely, for instance, on the actual cost to build and run the facility together with the costs paid by the facility clients. The first type of cost can be reasonably regarded as independent of the site chosen for location and it is not so relevant in the minimization process. In contrast, the costs paid by the facility clients determine a distribution of the potential users on the territory, significantly influencing the traffic flows through the transportation networks connecting the possible facility locations.
In a real urban context, the concept of "client" can be naturally replaced with the concept of "vehicle" or "transportation means", since it is impossible to ignore the presence of an underlying transportation network through which the i-th client reaches the j-th facility (Cascetta 1990, Gelmini 1986). According to this basic consideration, the model proposed in this paper describes the road network as an electric network, enabling to evaluate the incremental traffic due to the access to the service. It describes each lane of a road as an oriented link, characterized by (time-varying) parameters, such as the travel time at given unsaturated conditions. Road intersections are represented as nodes with inflows and outflows. The propagation delays due to the presence of traffic lights (Cantarella and Festa, 1998) and non homogeneous flows are also taken into account. Finally the effect of the location of a facility in a precise site over the surrounding extended area is modeled.
The evaluation of the "temporal" use of the urban services is nowadays a fertile investigation field (Bonfiglioli, 1997).

The proposed model allows urban operators to retrieve a significant number of related results, such as the influence area of each service site, the degree of utilization of each service, the induced traffic flow increment on each road, the total cost of any given service location. Moreover, other relevant information can be obtained: the minimum path matrix for all the nodes of the network; the distribution of traffic due to a given origin-destination matrix. For these reasons, the proposed model can be the basis of an interactive software tool to be used as a decision support system, for instance, by Urban Planning Experts or Public Administrators.

3.1 The model

- In order to define the model (Biancardi, De Lotto and Ferrara, 2000), the problem is to consider M facilities Sj, 1≤j≤M of a certain nature to be located in a urban context under the following assumptions:
- each facility has the assigned capacity to serve Cj clients per hour;
- the spatial distribution of the potential clients over the urban territory is known (more precisely, the spatial distribution is discretized into elementary units called "cells");
- each cell i, $1 \le i \le N$, contains $p_i(t)$ potential clients (time function valued in number of clients per hour);
- the i-th cell is centred in the i-th road intersection, $1 \le i \le N$, N being the total number of roads intersections of the urban transportation network considered.

The p_i-*th* client's choice of the facility to reach is dictated by the cost to access the facility. In the case of transportation by means of private vehicles, such a cost can be modelled as directly proportional to the vehicle travel time $t(i,j)$ from the i-th road intersection, from which the p_i-*th* client starts, to the j-th road intersection, where the facility is located. The global vehicle travel time is obviously the sum of the travel times associated with the links of the transportation network through which the client moves to reach the facility.

The urban transportation network can be represented by a graph with N nodes and a set of oriented branches $l(i,j)$, $i \le N$, $j \le N$, connecting the i-th node with the j-th node, with the travel direction from i to j. Each link is marked by a label which defines the time-varying travel time law on the link itself as a function of the traffic volume $n_{ij}(t)$. To further refine the model of the access to a facility the following aspects should be determinable:

- for any node, the time to access the nearest facility;
- the nodes "captured" by a facility, and the corresponding burden in terms of clients. This, in turn, allows one to identify the influence area of each facility delimited, on the nodes map, by the border lines connecting the nodes with associated longer access time;
- the number of clients reaching each facility;
- the induced traffic variation in each branch of the transportation network;
- the total cost for the users which is implied by the selected facility location: this quantity can be computed by summing up all the access times associated with the nodes multiplied for the number of clients arriving from each node.

3.2 The access time computation

To determine the quantities indicated above, it is necessary to compute the travel time corresponding to each link of the considered transportation network, making reference to the particular traffic conditions in the time interval of interest. The travel time is provided as a function of the traffic intensity, of the parameters which determine the traffic fluidity, and

the possible presence of traffic light (Cascetta 1990). Yet, in our case, the point is to evaluate the traffic variation due to the location of a certain facility in a certain area.
Then, the mentioned function can be linearly approximated by the tangent in the operation point in question. More precisely, given the regular traffic in the considered link, one can determine the corresponding travel time $t_0\ (i,j)$. Then, the travel time after that the facility has been located can be written as

$$t(i,\ j)\ =\ t_0(i,\ j)\ +\ R[n(i,\ j)\ -\ n_0(i,\ j)]\ =\ t_0(i,\ j)\ +\ R_{ij}\Delta n(i,\ j) \tag{1}$$

where $\Delta N(i,j)$ is the traffic intensity induced by the considered facility and

$$R_{ij}\ =\ \left.\frac{dt(i,\ j)}{dn(i,\ j)}\right|_0 \tag{2}$$

Clearly, in searching the optimal facility location, it is the average access time to be crucial rather than the access time of a single client. This is the reason why a macroscopic continuous-time model turns out to be the correct choice. Moreover, since there are plenty of efficient commercial tools for the analysis of electrical networks, it seems natural to depict an electrical equivalent of each link of the transportation network (Figure 1) in fact representing this latter as an electric network.

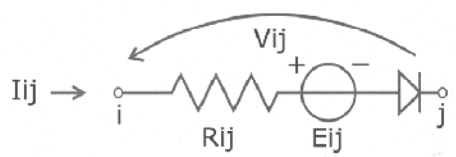


Fig. 1. The electric equivalent of a link of the transportation network

Note that, in the electric equivalent of a link, the presence of a diode guarantees that the current flows in a unique direction, and so does the traffic. The role of the voltage generator E_{ij} is to polarize the diode, thus representing the original travel time $t_0(i,j)$. The value of the resistor R_{ij} describes the dependence on the current I_{ij}, i.e., on the induced traffic intensity $\Delta n(i,j)$. Finally, the voltage V_{ij} represents the link travel time $t(i,j)$.
To determine the electric equivalent parameters corresponding to each link of the transportation network the following considerations can be made. The travel time, in seconds, along an urban road in regular traffic conditions is given by

$$t\ =\ \frac{3600\ \ \Lambda}{v}\ +\ t_s \tag{3}$$

where Λ is the length in Km of the path between two subsequent road intersections, v is the mean speed (Km/h) of the vehicles, t_s is the additional time due to the presence of a traffic light; t_s can be determined by well known empirical expressions which interpolate experimental data (Cascetta, 1990).

3.3 The model of the clients population

Relying on the electrical modeling equivalent, the clients' population can be modeled by means of ideal current generators which inject their current in the nodes of the

transportation network where the center of mass of each cell is located. The values of the currents (clients) are expressed in terms of the number of vehicles per hour (veh/h). They account for both the time distribution and the socioeconomic features of the clients population, this through a parameter related to the "appeal" of each facility. Note that with the term facility "appeal", author means the capability that a certain type of facility has to attract the clients' population. Data relevant to this capability can be acquired from national statistical studies centers (for instance, CENSIS). Generally, the available data provide the number of families attracted by the considered facility type, the number of persons for family unit, their distribution over the territory.

From the modeling point of view, each facility, regarded as incapacitated, is described by setting at a null potential the corresponding node where it is assumed to be located. The usage intensity of the considered facility is given by the sum of the currents entering such a node.

In alternative, as long as the facility has a limited capacity C_j (in clients/hour), the node where the facility is placed is not directly put to mass, but its connection to the null potential level is that depicted in Figure 2, where V_j is equal to zero until $I_j \leq C_j$, that is up to the facility saturation.

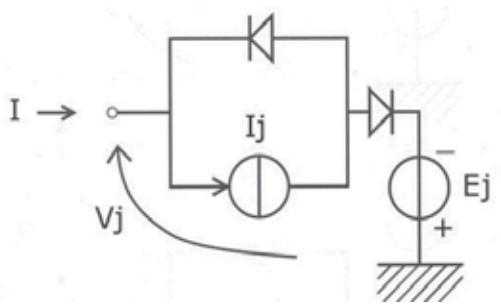


Fig. 2. The electric equivalent of a capacitated facility located in the j-th node

As soon as $I_j > C_j$, the diode is cut-off and V_j increases, practically re-distributing the clients among the other facilities, while keeping $I_j = C_j$. It is worth noting how this electrical effect resembles a waiting-in-queue time. More precisely, the queuing time of the j-th facility is modelled by the potential V_j.

Ej and the diode connected to it represent the facility appeal, measured as the extra-time the client accepts to spend to reach the service because of its appeal level, determined by statistical evaluation of fuzzy decision techniques (Lee and Zadeh, 1969) on clients' opinions.

3.4 Network solution

Given the parameters E_{ij} and R_{ij} for any link of the electric network representing the transportation network as a function of the operating point and of the dependence of the travel time on the traffic intensity, specified, in each node, the magnitude of the current generators modelling the clients population, and, finally, chosen a feasible location of the considered facilities, then, solving the network means to determine:

- the voltage at each node on the basis of which it is possible to quantify the access times, the border lines and the consequent partition of the network nodes which identifies the influence areas of each facility. Note that the facility access time is given by the

difference between the potential of the source node, i.e., the starting point, and the destination node, namely the node where the facility is located;
- the current in each link of the network which represents the traffic variation induced by the facility;
- the current in the node where the facility is located which describes the degree of utilization of the facility.
- It is worth noting that the network solution, relying on the analogy between voltage at the nodes and travel times, and on the analogy between currents in the links and induced traffic, leads to the determination of the minimum of the cost function *C* (Di Barba and Savini, 1996). Indeed, according to (Maxwell 1892), in a circuit made up by resistors, independent voltage and current sources, the currents tend to reach a distribution such that the dissipated power is minimum.

It is interesting to note that the total cost gives the number of vehicles the chosen location induces in the road network, so its minimization means the minimization of induced traffic pollution.

The advantage of the modeling analogy pursued is mainly tied to the possibility of exploiting the huge capabilities of commercial electrical networks analyzers to determine the relevant parameters characterizing the behavior of the clients' population with respect to a certain choice of the facilities location.

To this end, one of the most common tools is SPICE, which can be regarded as a standard all over the world to analyze and simulate electric networks (Nagle 1975).

The problem of solving the electric networks which accounts for the effects on the underlying transportation network of the facility location choice belongs to the class of problems that can be easily dealt with by SPICE. Moreover, the computing time is such that even a general-purpose PC can be used. Once that the features of the considered network have been acquired by SPICE, it computes the voltage at each node and the currents in the links.

4. The system ULISSE

The main purpose of ULISSE consists of creating an unmediated interaction with context data and of supporting the urban planner in the investigation of planning choice effects.

The first difficulty is that no circuit simulation program has a user interface that can be adapted to the urban context; the only option would be to translate the viability graph into a net-list describing the whole graph element by element, but this operation would lose any geographical information defining the location of the graph nodes.

On the other hand, Geographical Information Systems (GIS) cannot be of any help either, since the problem lies in the way circuit parameters are managed and related to geographical information. So the choice author made has been of developing an ad-hoc tool that could interact with SPICE to process the traffic network data, yet let urban planners investigate their alternatives in a totally graphical way.

A typical working session would begin with the designer setting up the planning scenario as follows:
- read a map of the town or the area to be studied;
- mark on the map, and enter the respective parameters of, facility locations, client cells (nodes that represent the population of a neighborhood), and transportation network additional nodes;

- define the connection arcs and enter their quantified values;
- run the simulation;
- get the results.

The session would then explore the planning opportunities in a three-step process: updating of project parameters, simulation run and result analysis.

It is clear that, while the simulation is performed by SPICE, all the tasks concerning the initialisation and the update of scenario parameters and everything that deals with the analysis of results are carried out by interacting with the user: it is precisely on these tasks that author focused trying to provide planners with a system that would not hinder their creativity, but rather that would foster it actively.

4.1 Effective viability graph input

The first step a planner takes, unless retrieving a previously saved project, is to set a cartographic map as the project graphic reference; all the other elements belonging the scenario will be placed according to the project map, as showed in Figure 3. ULISSE lets planners dimension the map by using a calibration command and then by defining a segment of known length.

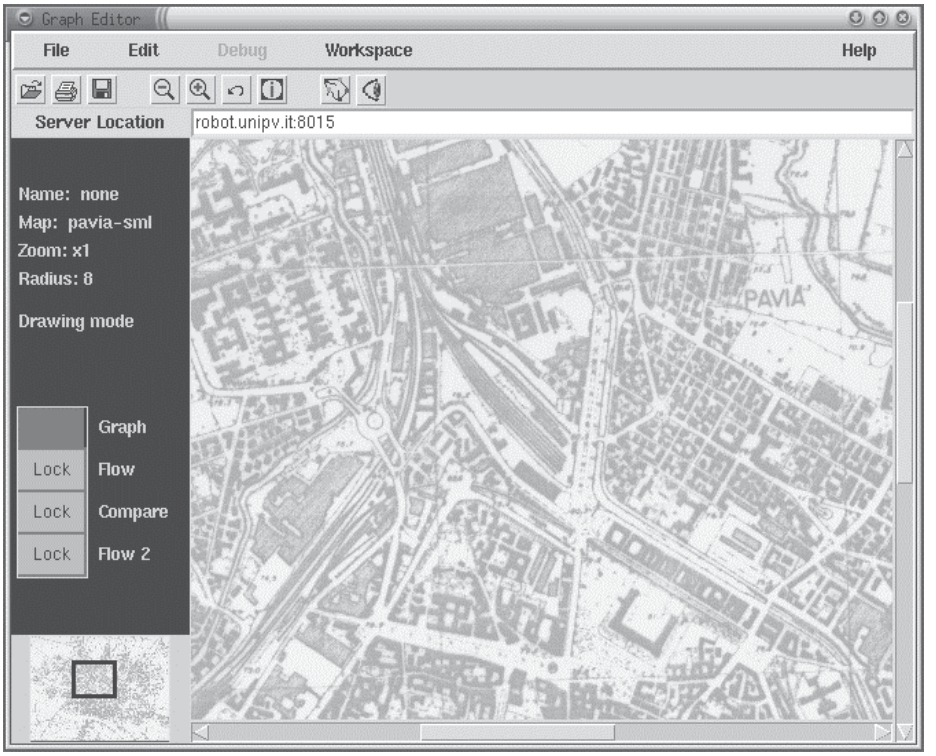


Fig. 3. Screen shot of ULISSE with the map of Pavia loaded

The next steps are the introduction of traffic junctions and branches (nodes and arcs of the urban graph), as shown in Figure 4. Actually, author established that the arcs of the viability graph could only be drawn by connecting existing nodes.

This operational choice in the ULISSE GUI was made in order to speed up all the project graph initialization: ULISSE users are thus free from a mandatory use of the toolbar to switch between arc and node input modes.
Once all the graph data are entered, the network is ready to be simulated and analyzed. Anyway, it is always possible to interact with the scenario by changing any of the graph parameters or by adding or removing nodes and arcs at any time.

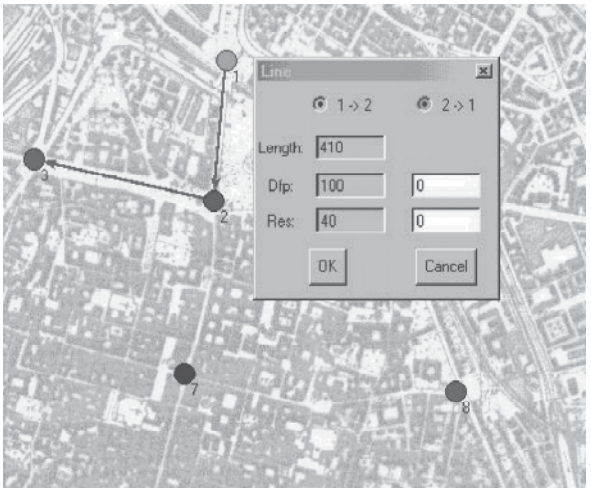


Fig. 4. Net building

5. Result and applications

ULISSE allows urban operators to retrieve a significant number of results, such as the influence area of each service site, the degree of utilization of each service, the induced traffic flow increment on each road, the total cost of any given service location.
Moreover, other relevant information can be obtained: the minimum path matrix for all the nodes of the network; the distribution of traffic due to a given origin-destination matrix.
These are typical results of an accessibility model, and ULISSE permits also to face many other urban subjects such as environmental sustainability of settled scenarios, minimum level of efficiency of public mobility services, localisation of emergency services. Applications of the system given in this paper concern the city of Pavia, in which interesting problems regard the reuse of ex-industrial areas and mobility in the centre of the city - related to its specific morphology and to the distribution of services, which is concentrated in this part of the town.
With reference to official traffic data (Urban Mobility Plan, Municipality of Pavia, 2008), in order to face different matters, the examples pertain two mobility nets: the first regarding the whole city (with 90 nodes), the latter focused on the centre (62 nodes).

5.1 Environmental impact check

The aim of this application is to verify the pollution effects due to a settled distribution of services located in two ex-industrial areas, now apt to an urban renewal: National Railway goods yard (node 23) and SNIA Viscosa (node 82) as shown in Figure 5.

In these areas residential buildings, public services and commercial ones will take place; ULISSE will check the environmental impact of the additional traffic driven by large retailers.
With the optimal distribution of clients, ULISSE discovers critical pollution branches. Pollution index is measured by the traffic flow in a given branch in cars/sec multiplied by its path covered time. The model gives this quantity by the electrical power dissipated in the chosen branch: P = V·I.

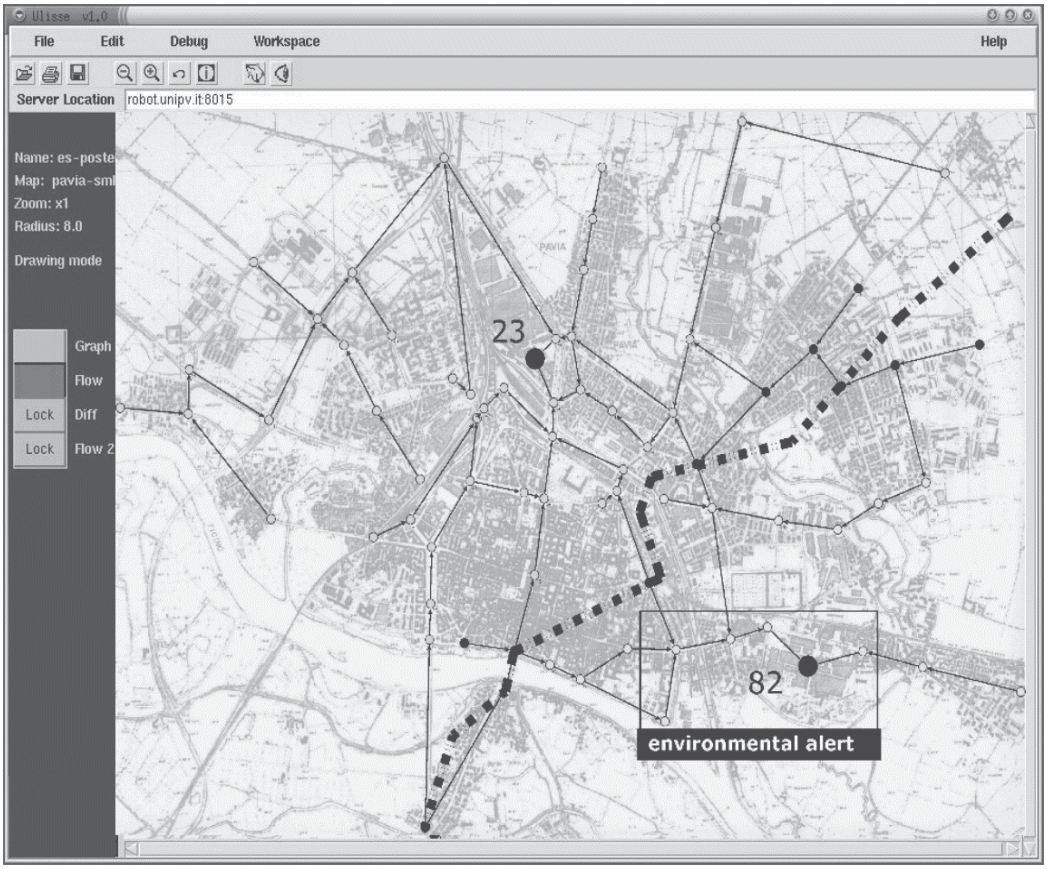


Fig. 5. Net with environment alert around node 82, and the influence zones divided by the ridge line

Figure 5 shows the plan of the city with the considered mobility net, the ex-industrial areas (site of new large retailers) and the distribution of influence areas for each service. The ridge line shows how node 23 receives a larger influence area: this result depends on existing traffic, which is usually dense in the south-east of the city (crossover traffic to east).
The system graphically signs a critical area in the net around node 82 (Figure 5 and 6):

- branch 38-65 (363 meters): pollution index equal to 30 cars – with a density of one car every 12 meters – situation of very dense traffic;
- branch 59-65 (760 meters): pollution index equal to 38 cars – one car every 20 meters;
- branch 65-66 (257 meters): pollution index equal to 27 cars – one car every 9.5 meters, close to complete saturation evaluated in one car every 7 meters (Figure 6).

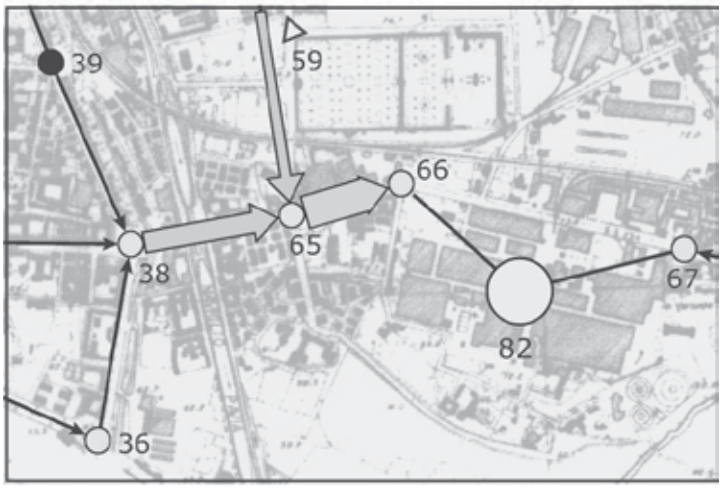


Fig. 6. Thematic zoom on critical branches

Considering the existing traffic, it is easy to have a holdup in branch 65-66, and these traffic conditions last for 25-30 minutes. Environmental consequences of the specific localisation choice are remarkable.

With this results, planners should find different solutions such as: modify of streets fluxes introducing a one-way branches; build new mobility branches (i.e. in the southern area); consider different localisation possibilities in the east part of the city.

5.2 Efficiency of public mobility services

In Pavia, mobility in the centre is critical especially considering the high number of services. Another application of the system is about the multi-modal mobility network, regarding private mobility, public mobility, pedestrians and bicycles. A recent strategy for traffic calming in historical cities is to define pedestrian areas in the city centre, with private movement permission only for inhabitants. Pavia is an archaeological site, and a subway is forbidden: the only feasible public mobility service is on the soil (bus). To reach central services in a reasonable time, bus net must be efficient and needs inter-modal nodes, characterized by parking areas and stations. In ULISSE, multi-modal nodes are considered as services images: destination nodes with positive voltage value (representing the time to reach the real service by bus). Here, this value corresponds to the efficiency of public lines to be minimized.

Taking into account a new net focused on the centre of the city, in which 9 external nodes connect to the remaining real net, 3 services are considered. Two in the centre (S1 and S2) and one further external (S3).

At first step 2 multimodal nodes are assumed, with an equivalent access time equal to 600 seconds each (waits + travels); the resulting total cost is equal to 74.564 second. If the whole net, with the same services, could be used by private cars the result would be equal to 55.545 seconds, and this quantity is considered as reference.

Considering 4 multimodal nodes (b1, b2, b3, b4), with an equivalent access time equal to 600 seconds (wait + travel) the resulting total cost is equal to 69.873 seconds.

Modifying the equivalent access time for each single multimodal node, and verifying the global cost, an acceptable result (global cost close to 60.000 seconds) is for public transport

working at 480 seconds (wait + travel) with 4 multimodal nodes, located as in Figure 7, with influence areas as shown in Figure 8.

Fig. 7. New net with multimodal nodes (bi) and services (Si)

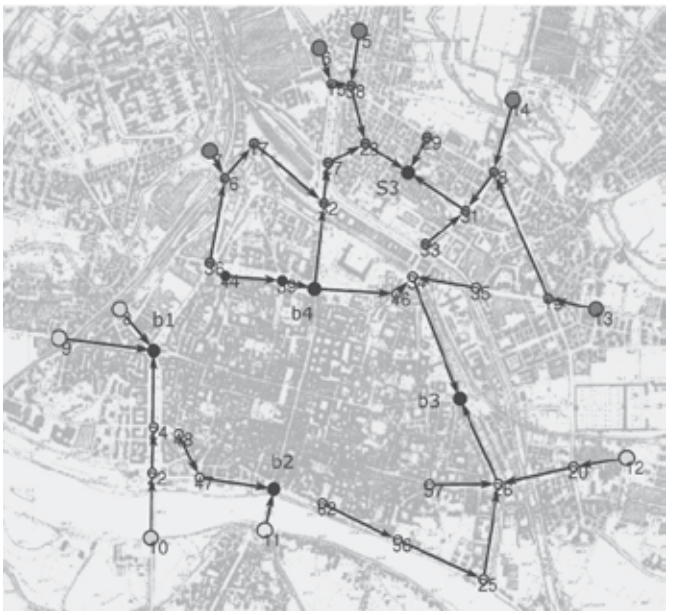


Fig. 8. Influence areas with multimodal mobility net

5.3 Emergency services localization

To obtain the best location for emergency services, such as ambulance basis, the problem is minimizing the access time from each starting point to every possible destination. Compared with the previous nets, it means to change each node from client to service and vice versa. Because of the considered model (electrical analogy), it is possible to consider the net as perfectly symmetrical; in traffic meaning, authors have to consider only existing double sense streets, with similar characteristics for both directions. This simplification is perfectly plausible considering the usual pace of emergency means (acoustic and visual signals ease streets vacation).

Fig. 9. One single emergency base

Fig. 10. Two emergency basis

To test the net, in the first example it is considered one single "ambulance base", located in node E in the centre of the city (Figure 9). The access time is considered on the most distant nodes (net perimeter). In this first case, the maximum access time is at node 14, with 492 seconds, while for node 5 it is 468 seconds.
Then two emergency basis (E1 and E2) have been located (Figure 10). The maximum access time is at node 12, with 358,5 seconds.
Locating 3 basis (E1, E2 and E3), the system gives a general lower access time but, even moving the basis in every suitable place, for the further node it is impossible to decrease the access time lower than 302 seconds. In view of reaching a better result, four bases are situated considering all the possible locations. Comparing the different possibilities, with the best location choice, the maximum access time to node 13 with 220 seconds.

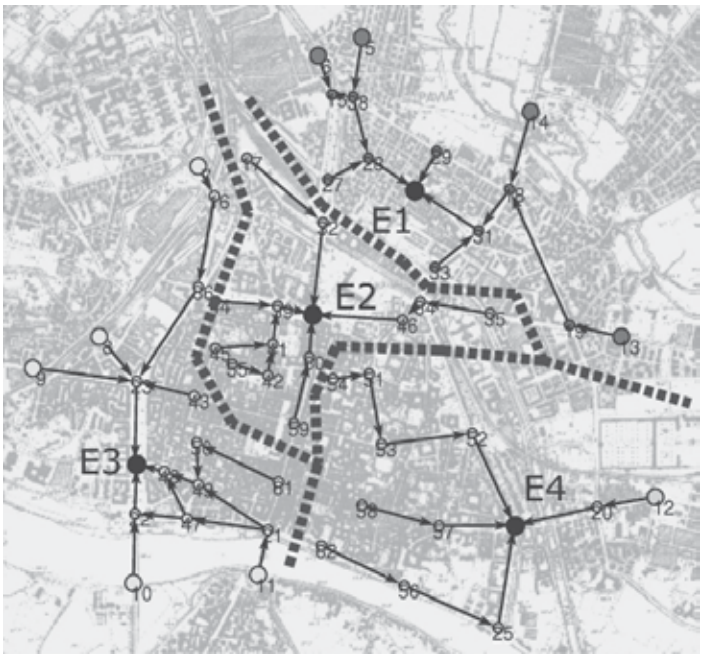


Fig. 11. Four emergency basis - influence areas and ridge lines

6. Discussion

The proposed model is an example of specific technical tools that planners can use to get fast and cheap outputs of localisation hypotheses during the first phases of planning processes. Urban planning process is extremely complex and when planners define the main structure of the plan, they need to verify the order of magnitude of the mass of parameters that are involved in the process. This operation is functional if planners can execute it by themselves. More specific analyses are required in other phases, after the main idea is defined.
ULISSE offers a simple way to calculate the impact on traffic due to location decisions (as well as localisation opportunities with given accessibility requirements) and can be considered as a tool of a comprehensive Decision Support System for urban planning decisions about location problems.

In the paper, urban services localisation is considered in terms of users' access time and of clients' population movement, taking into account the underlying transportation network connecting the various facilities.
Since ULISSE provides a series of standard location like results (such as the influence area of each service site, the degree of utilization of each service, the total cost of any given service location, the minimum path matrix for all the nodes of the network, the distribution of traffic due to a given origin-destination matrix) a significant number of related marks can be determined relying on the proposed model with a friendly interface.
The model has been tested in cities like Pavia, which is a medium city in Italian context; (basis dimensional data are: 70.000 inhabitants, 60 sqKm of Municipality extension, almost 15 sqKm of urbanized area).
Pavia road system is quite simple because there are:

- a clear scheme composed by the main inner circle (ancient walls location) and radial axes;
- a small number of streets (7) that connect the urban area to the external land.

The application of ULISSE in bigger or more complex contexts could provide less satisfactory results.
In fact, ULISSE needs information regarding traffic conditions in every branch of the considered net at the t_0 instant. In a bigger city (or road net), data collecting could be very expensive.
Considering the specific target of the system, which is to give to planner the order of magnitude of accessibility problems, the use of approximate data (taken from statistical elaborations) is acceptable only in limited contexts.
In ULISSE every node and every branch of the net must be entered manually, and for big nets the use of the system is expensive in term of data entry time.
Application of ULISSE in more complex nets could be affected by its simplified scheme of road intersections, which do not consider, in example, specific turns. In particular cases, road intersections can be considered as sub-nets, but in this way the intricacy of the net (and of relative data requirement) increases.
Further steps of development of ULISSE consist on giving to the user the possibility to choose different simulation models, i.e. dynamic models, with the same graphic user interface.
Actual traffic dynamic models require a greater quantity of data in comparison to the static model here presented, but surely can provide more precise outputs and consequently suggest a larger number of possible applications.

7. Acknowledgments

At the development of ULISSE were involved Prof. Antonella Ferrara (actually at the University of Pavia) who worked on the model optimization and Prof. Alberto Biancardi (actually at the Cornell University) who worked on the graphic user interface.
I thank them very much for their fundamental help.

8. References

Alonso, W. (1964). *Location and land use: towards a general theory of land rent*, Harvard University Press, Cambridge, Mass.

Benevolo L., (1963). *Le origini dell'urbanistica moderna*. Universale Laterza, Bari, Italy. In Italian.

Bertuglia, C. S., Staricco, L. (2000). *Complessità, autoorganizzazione, città,* Franco Angeli, Milan. In Italian.

Biancardi A., De Lotto R., Ferrara A. (2000). A Multi-modal Transport Simulation Tool to solve Urban Location Problems, *Simulation in Industry*, SCS, Hamburg, 437-442

Bonfiglioli, S. (1997). "Urban Time Policies in Italy, an Overview of Time Oriented Research, *Transfer. European Review of Labour and Research*, Bruxelles, n. 4, 700-722

Cantarella, G. E., Festa, D. C. (1998). *Modelli e Metodi per l'Ingegneria del Traffico*. Franco Angeli, Milan. In Italian.

Carlstein, T., Parkes, D., Thrift, N. (1978). *Timing space and spacing time*, 3rd vol., Arnold, London.

Cascetta, E. (1990). *Metodi Quantitativi per la Pianificazione dei Sistemi di Trasporto.* CEDAM, Padova. In Italian.

Daskin, M. S. (1995). *Network and Discrete Location, Models, Algorithms and Applications*, Wiley Interscience, New York.

De Lotto, R., Ferrara, A. (2001). An analog differential model to solve location problems in urban planning, *Ricerca Operativa*, Franco Angeli, Milan, vol. 31 n.97, 3-22.

De Lotto, R., Ferrara, A , (2002). A transportation network model as a tool to solve urban planning location problems, *Mathematical and Computer Modelling of Dynamical Systems,* Swets Zeitlinger Publisher, Lisse, Netherlands, vol. 8, n°3, 313-332.

Dell'Amico, M., Maffioli, F., Martello, S. (1997). *Annotated Bibliographies in Combinatorial Optimization* , John Wiley and Sons, New York.

Di Barba, P., Savini, A. (1996). An optimization approach to the analysis and synthesis of circuits and fields. *Proceedings of 3rd International Conference on Computation in Electromagnetics (CEM 96)*, Bath, n. 420, 370-375.

Donato, F., Lucchi Basili, L. (1996). *L' ordine nascosto dell'organizzazione urbana. Un'applicazione della geometria frattale e della teoria dei sistemi auto-organizzati alla dimensione spaziale*, Franco Angeli, Milan. In Italian.

Drezner, Z. (1995). *Facility Location, A Survey of Application and Methods*, Springer-Verlag, New-York.

Gabellini, P. (2001). *Tecniche urbanistiche*. Carocci, Milan. In Italian

Gelmini, P. (1986). *Modelli Urbanistici di Distribuzione*. CLUP, Milan. In Italian.

Hoffmann, R. (1997). Time in the city – a new action field for trade unions, *Transfer. European Review of Labour and Research*, Bruxelles, n. 4, 775-785.

Körkel, M. (1989). On the exact solution of large-scale simple plant location problems, *European Journal of Operational Research*, n. 39,157-173

Lee, E. T., Zadeh, L. A. (1969) Note on fuzzy languages. *Information Science*, 1, 421-434.

Maxwell. J. C. (1892*). A Treatise on Electricity and Magnetism*, Oxford Press, Oxford

Mazza, L. (2002). *Piano, progetto, strategie*, Franco Angeli, Milan. In Italian

McLoughlin J. B. (1969). *Urban and Regional Planning: A System Approach*, Faber and Faber, London

Nagel, L. W. (1975). *SPICE: a computer program to simulate semiconductor circuits.* ERL Memo M520, Electronics Research Laboratory, University of California, Berkeley, CA.

Rabino, G.A. (2005) *Processi Decisionali e Territorio nella Simulazione Multi-agente*. Esculapio, Bologna. In Italian.

Vicini D. (1988). *Pavia: materiali di storia urbana: il progetto edilizio 1840-1940*, Pavia Municipality, Italy. In Italian.

24

Design and Development of a Compound DSS for Laboratory Research

Tomáš Hujer
University of West Bohemia in Pilsen
Czech Republic

1. Introduction

Using of decision support systems is today far away from being only the domain of top business management. DSSs are successfully applied in many areas of human activities, from traditional finance, financial forecasting and financial management, through clinical medicine, pharmacy, agronomy, metallurgy, logistics and transportation, to maintenance of machinery and equipment.

Despite this, the use of decision support systems in the domain of laboratory research is still relatively unexplored area. The main idea behind the application of DSS in this particular domain is increasing the quality and shortening the duration of research, together with reducing costs. To achieve these objectives, making the right decisions at the right time using the right information is needed. Unfortunately, the disadvantage of decision support in the field of laboratory research is mainly the lack of historical data. The rules for decision-making are still nascent during the research. This makes the issue of applying DSS for laboratory research very interesting.

It is obvious that requirements for computer support of laboratory research will vary from case to case, sometimes even substantially. On the other hand, there is a characteristic common to all laboratory research. The laboratory research consists of series of tests and measurements which generates data and knowledge as their outputs. To make a research effective, it is necessary to apply an appropriate process control to diagnostics, as well as knowledge acquisition techniques and knowledge management tools. Moreover, knowledge is very often hidden in the relationships between measured data and has to be discovered by using sophisticated techniques, such as Artificial Intelligence.

There are several options for building DSS application. This chapter is focused on in-house development as the best way to develop DSS application with maximal possible compliance with user's demands and requirements. Especially evolutional prototyping enables rapid development and deployment of the system features and functions according to the actual user's requirements. On the other hand, in-house development puts certain requirements on IT skills, which may be an intractable obstacle in some cases.

The objectives of this work are not to describe the universal, ready to use DSS, but to reveal possibilities, means and ways, to describe the methodology of design and in-house development of DSS for laboratory research with the most possible fit to user's requirements.

2. Types of DSS suitable for laboratory research

There are several classifications and taxonomies of DSS applications. A brief overview of DSS classifications brings French (French et al., 2009). For this work, the AIS SIGDSS[1] classification is used. It is today the most common and widespread DSS classification, popular with many authors, such as Power (Power, 2000) or Turban (Turban et al., 2008). This classification divided DSS into five main categories as follows:

- **Data-driven DSS**, which are primarily based on the data and their transformation into information,
- **Model-driven DSS**, which puts the main emphasis on the use of simulation and optimization models,
- **Knowledge-driven DSS**, characterized by the use of knowledge technologies to meet the specific needs of decision-making process,
- **Document-driven DSS**, that helps users acquire and process unstructured documents and web pages, and
- **Communications-driven and group DSS**, which includes all systems using communication technologies to support collaboration of user groups.

Of course, all listed categories can be combined to create compound or hybrid systems. For laboratory research purposes, the combination of data- and knowledge-driven DSS seem to be the best solution. Laboratory research is based on diagnostics. Important attributes with which the diagnostic processes work are data, information and knowledge. Knowledge transforms data into information and information (diagnosis) is the output of diagnostic process (Tupa, 2008). The diagnostic process is schematically illustrated in Figure 1.

Fig. 1. Process view of diagnostics (Tupa, 2008)

As already mentioned, Data- and Knowledge-driven decision support systems are the most appropriate types of DSS for laboratory research. The next two subchapters are focused on definition, characteristics, and the key features of these two types.

2.1 Data-driven DSS

A Data-driven decision support system is defined as an interactive computer-based system that helps decision-makers use large database. Data-driven DSS primarily rely on data and their processing into information, along with the presentation of this information to a decision maker (Turban et al., 2008). The main goal of such systems is to help users transform data into information and knowledge. Users of these systems can perform

[1] Association for Information Systems Special Interest Group on Decision Support System

unplanned or ad hoc analyses and requests for data, process data to identify facts and draw conclusions about the data patterns and trend. Data-driven DSS help users retrieve, display, and analyze historical data.

This broad category of DSS help users "drill down" for more detailed information, "drill up" to see a broader, more summarized view, and "slice and dice" to change dimensions they are viewing. The results of "drilling" and "slicing and dicing" are presented in tables and charts (Power, 2000). With this category of DSS are mainly associated two technologies – Data Warehousing and Online Analytical Processing (OLAP).

Data warehouse is defined as subject-oriented, integrated, time-variant, non-volatile collection of data in support of user's decision making process. Subject-oriented means it is focused on subjects related to examined activity. Integrated means the data are stored in a consistent format through use of naming conventions, domain constraints, physical attributes and measurements. Time-variant refers to associating data with specific points in time. Finally, non-volatile means the data do not change once they are stored for decision support (Power, 2000).

OLAP and multidimensional analysis refers to software for processing multidimensional data. Although the data in data warehouse are in multidimensional form, OLAP software can create various view and more dimensional representations of data. OLAP usually includes "drill down" and "drill up" capabilities. This software provides fast, consistent, and interactive access to shared multidimensional data.

The data-driven DSS architecture involves data store, data extraction and filtering component, end user query tool, and end user analysis and presentation tool. The data store consists of database or databases built using relational, multidimensional, or both database management systems. The data in data store are summarized and arranged in structures optimized for analysis and fast retrieval of data. The data extraction and filtering component is used for extract and validate the data taken from so called operational database or from external data sources. It selects the relevant records and adds them to the data store in an appropriate format. The end user query, analysis and presentation tools help users create queries, perform calculations, and select the most appropriate presentation form. The query and presentation tools are the front-end to the DSS.

Data-driven DSS are usually developed using general development approaches called System Development Life Cycle (SDLC) and Rapid Prototyping (see Section 4), depending on the size of the resulting system.

2.2 Knowledge-driven DSS

Knowledge-driven DSS is overlapping term for the decision-making support systems using artificial intelligence technologies. These systems are usually built using the expert system shells and data mining or knowledge discovery tools. Knowledge-driven DSS era interactive programs that made recommendations based on human knowledge. This category of DSS helps users in problem solving, uses knowledge stored as rules, frames or likelihood information. In addition, these systems may have capabilities to discover, describe, and predict knowledge hidden in data relations and patterns.

According to Turban (Turban et al., 2008), expert systems (ES) are computer-based system that use expert knowledge to attain high level decision performance in a narrow problem domain. ES asks questions and reasons with the knowledge stored as part of the program about a specialized subject. This type of program attempts to solve a problem or give advice (Power, 2000).

ES consists of three elements – knowledge base, inference engine, and user interface. Knowledge base contains the relevant knowledge necessary for understanding, formulating and solving problem. Typically, it includes two elements – facts that represent the theory of the problem area, and rules or heuristics that use knowledge to solve specific problem. Inference engine is the control structure or the rule interpreter of ES. It is a computer program that derives answers from knowledge base and formulates conclusions. Inference engine is a special case of reasoning engine, which can use more general methods of reasoning. User interface is a language processor that provides user-friendly, problem-oriented communication between the users and the expert system.

The aim of data mining (DM) is to make sense of large amounts of mostly unsupervised data in some domain (Cios et al., 2007). Data mining techniques can help users discover hidden relationships and patterns in data. They can be used either for hypothesis testing or for knowledge discovery. According to Power (Power, 2000), there are two main kinds of models in data mining – predictive and descriptive. Predictive models can be used to forecast explicit values, based on patterns determined from known results. Descriptive models describe patterns in existing data, and are generally used to create meaningful data subgroups.

Data mining software may use one or more of several DM techniques. These technique and DM tools can be classified based on the data structure and used algorithm. The most common techniques are:

- **Statistical methods**, such as regression, correlations, or cluster analysis,
- **Decision trees**, that break down problems into increasingly discrete subsets by working from generalization to increasingly more specific information,
- **Case Based Reasoning**, that uses historical cases to recognize patterns (see Section 3.2),
- **Intelligent agents**, that retrieve information from (especially) external databases, and are typically used for web-based data mining,
- **Genetic algorithms**, that seek to define new and better solution using optimization similar to linear programming,
- **Neural computing**, that uses artificial neural networks (ANN) to examine historical data for patterns and applying them to classification or prediction of data relationships (see Section 3.4),
- **Other tools**, such as rule indication and data visualization, fuzzy query and analysis, etc.

Knowledge-driven DSS are usually built using several proposed Rapid Prototyping approaches.

3. Diagnostic DSS

Diagnostic decision support systems are mainly associated with the domain of clinical medicine. They are developed since the early seventies, and are designed to provide expert support in diagnosis, treatment of disease, patient assessment and prevention. Between the medical and technical diagnosis is an obvious similarity. In technical diagnosis, the examined subject is also analyzed to obtain a diagnosis. On its basis, corrective and preventive actions can be proposed and taken, and examined subject's condition can be monitored, evaluated and predicted.

Although the use of diagnostic decision support systems overlaps with other fields, such as maintenance planning and management (Liu & Li, 2007), or the prediction of product life cycle (Lolas & Olatunbosum, 2008)(Li & Yeh, 2008), the vast majority of available

publications dealing with application in clinical medicine. Vikram and Karjodkar (Vikram & Karjodkar, 2009) briefly summarize the history of so called clinical decision support systems (CDSS) development. According to their findings, there are four main types of CDSS, which are based on:

- **Rule Based Reasoning** (RBR) – use of the principle of cause and effect (if-then),
- **Case Based Reasoning** (CBR) – decision making based on the principle of analogy with already resolved cases,
- **Bayesian believe networks** (BBN) – the use of probability theory,
- **Artificial neural networks** (ANN) – computational model inspired by the structure and functional aspects of biological neural networks.

Apart from these, there are few examples of the use of uncommon techniques, such as heuristic algorithms, fuzzy logic (Lingaard et al., 2007), or game theory (Lin et al., 2009). The following subchapters briefly describe the four main techniques used in diagnostic DSS and provide examples of use and suitability of these technologies for their use in laboratory research.

3.1 Rule based reasoning (RBR)

The rule based reasoning uses notation of rules in the "if-then-else" form. This notation can be extended to define the probability of suitability of the proposed actions. Rule-based reasoning is mainly used by expert systems that analyze the base of facts and apply the appropriate rules on the solved situation. The main disadvantage of RBR is laborious and time-consuming creation of a quality knowledge base.

Kumar (Kumar et al., 2009) describes a hybrid approach, combining case and rule based reasoning for branch independent CDSS for the intensive care unit (ICU). DSS that use rule based reasoning are usually limited to use in a specific area, such as cancer, poisoning, cardiology, etc. It is significantly limiting for multidisciplinary applications such as the DSS for ICU. The rigidity of the system was eliminated by combination with case based reasoning. CBR has been chosen as the main decision support technique. CBR uses RBR subsystem with knowledge base containing common rules for all medical disciplines needed at the intensive care unit.

In the area of research, the use of rule based reasoning seems to be unusable. The process of research is characterized by exploring and revealing relationships and rules within the survey data, therefore it is difficult, or impossible, to define the rules beforehand. Especially in the early stages of research is the use of rule based reasoning as a tool for decision support absolutely inconceivable.

3.2 Case based reasoning (CBR)

Case based reasoning is technique of computer-based decision-making which uses the principle of analogy with already resolved cases. It is based on the premise that the newly solved problems are often similar to previously solved cases, and therefore the previous solution can be used in current situation. The fundament of CBR is case repository, so called case library. It contains a number of previously solved cases, which are used for decision support. CBR is often used in medical DSS. One possible reason is that the reasoning based on previous cases is psychologically more easily acceptable than reasoning based on rule model (Turban et al., 2008).

Ting (Ting et al., 2010) describes a DSS integrating case based reasoning and association rules mining for decision support in prescribing of medications. According to him, CBR,

unlike Bayesian networks and ANN, does not have a tendency to generalize too much, which results in superior accuracy when proposing a solution derived from the memorized cases. Association rules mining is a technique used for extraction of significant correlations of frequent patterns, clusters, or causal structures among database items.
Zhuang (Zhuang et al., 2009) describes a new methodology of integrating data mining and CBR for intelligent decision support for pathology ordering. The purpose of the integration of data mining and CBR is gathering of knowledge from historical data using data mining, and retrieve and use these data for decision making support.
As the knowledge base for rule based reasoning, case library is a fundamental building block for decision support systems that use case based reasoning. For similar reasons as for rule based reasoning, CBR technique is not eligible for application in laboratory research.

3.3 Bayesian believe networks (BBN)

Bayesian believe network is a directed acyclic graph $G = (V, E, P)$, where V is a set of nodes representing random variables, E is a set of edges representing the relationships and dependencies between these variables, and P represents associated probability distributions on those variables. It is a graphics model capable of representing the relationships between variables in a problem domain.
In other words, BBN is directed graphical model where an edge from A to B can be informally interpreted as indicating that A "causes" B. The example of a simple BBN is in Figure 2. Nodes represent binary random variables. The event "grass is wet" (W=true) has two possible causes: either the water sprinkler is on (S=true) or it is raining (R=true). The strength of this relationship is shown in the table below W; this is called W's conditional probability distribution (NNMI Lab., 2007).

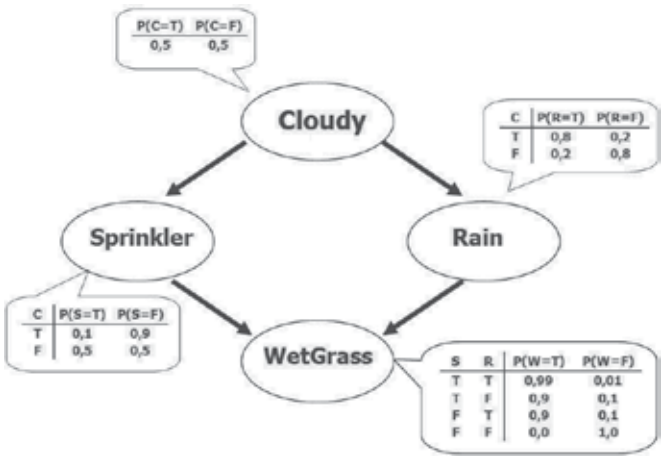


Fig. 2. A simple Bayesian network (NNMI Lab., 2007)

Bayesian networks represent good combination of the data and prior expert knowledge. Relationships between the variables in a problem domain can be interpreted both causally and probabilistically. BBN is able to cope with the situations where some data are missing, and unlike the rule based systems allow the capture of a broader context.
The application of Bayesian probability theory in diagnostic tasks addresses Lindgaard (Lindgaard et al., 2007). He claims that the Bayesian algorithm is a suitable alternative for

the diagnosis of disease, and thus to provide effective decision support in medical diagnostic systems. The result of Bayesian analysis is a set of hypotheses associated with the probability distribution. To develop decision rules, these probabilities are combined with information about the nature of possible decisions, their significance and relevance. Probabilities of Bayesian network nodes are then reviewed and further refine in each of the next iteration with new information set.

Liu and Li (Liu & Li, 2007) bring the example of using diagnostic DSS outside the field of clinical medicine. They used Bayesian networks to build a DSS for machinery maintenance, using BNN's suitability for applications in fault diagnostics. Described DSS supports the strategy of proactive maintenance based on monitoring and diagnosis of machines, and forecasting and prevention of disorders.

Using of Bayesian believe networks for the decision support poses two major problems. The first is the need of mastering the Bayesian probability theory, which means managing relatively large and robust mathematical apparatus. The second is the considerable computation complexity of algorithms for learning BBN from data, respectively difficult inference in large models.

Although Bayesian networks seem to be an appropriate technology for decision support in the field of laboratory research, above mentioned difficulties make the development of DSS applications only through own forces very difficult or even impossible for many users.

3.4 Artificial neural networks (ANN)

Artificial neural network is a computational model derived from the way of information processing performed by human brain. ANN consists of simple interconnected elements for data processing - artificial neurons. These elements process the data in parallel and collectively, in a similar way as the biological neurons. Artificial neural networks have some desirable properties similar to biological neural networks, such as learning ability, self-organization, and fault tolerance (Turban et al., 2008).

The basic element of the artificial neural network is an artificial neuron. Artificial neuron is a computation unit with inputs, outputs, internal states and parameters. This unit processes the input data (signals) and generates appropriate outputs. There are several types of artificial neurons, which vary according to the type of neural network. Generic type of artificial neuron, so called formal neuron, is shown in Figure 3.

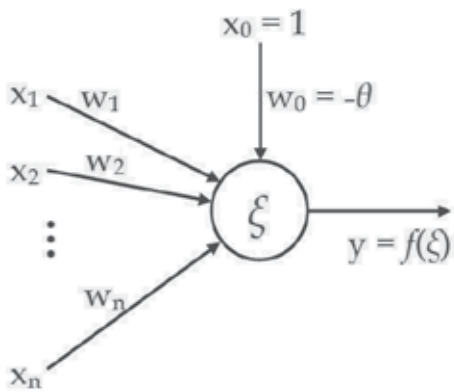


Fig. 3. Formal neuron

Formal neuron consists of several inputs ($x_1,\ldots, x_n$) and their connection weights ($w_1,\ldots, w_n$), formal input (x_0) and its connection weight, so called bias or threshold (w_0), neuron's own body, where the computation of the output is made, and one output (y), which can be further branching. The neuron function itself is divided into two steps. First, postsynaptic potential (ξ), i.e. weighted sum of all inputs, including formal input, is calculated. In the second step, so called activation function is applied to postsynaptic potential and its result is the value of the neuron output.

The connection weight is a key element of every ANN. It expresses the relative importance of each neuron input or, in other words, the degree of influence of the particular input to the output. Connection weight represents storage of patterns learned from the input data. The ability of learning lies exactly in the repeated refinement of the connection weights.

Because of their ability to learn and generalization, artificial neural networks are used in many applications of prediction and data classification. Aburas (Aburas et al., 2010) use neural networks to predict the incidence of confirmed cases of dengue fever. This prediction is based on the observations of real parameters, such as average temperature, average relative humidity, total rainfall, and the number of reported cases of dengue fever as a response to these parameters.

Faisal, Ibrahim and Taib (Faisal et al., 2010) also deal with the issue of dengue fever disease. They proposed a non-invasive technique to predict the health risks to ill patients via combination of self-organizing maps and multilayer feed-forward neural networks. Combining these techniques, they achieve 70% accuracy of forecasts.

Gil et al. (Gil et al., 2009) describe the use of artificial neural networks in the diagnosis of urological disorders. To suppress the main neural networks drawbacks, so called over-learning or over-fitting, they use a combination of three different ANN architectures, two unsupervised and one supervised. This combination has provided decision support with verified accuracy of almost 90%.

Both, Faisal (Faisal et al., 2010) and Gil (Gil et al., 2009), also mentioned the possibility of increasing the accuracy of their systems by combining artificial neural networks with fuzzy inference techniques. This approach uses Kannappan (Kannappan et al., 2010) for design the system for prediction of autistic disorders using fuzzy cognitive maps with nonlinear Hebb learning algorithm. Fuzzy cognitive maps combine the strengths and virtues of fuzzy logic and neural networks.

As mentioned earlier, artificial neural networks are often used for prediction, or to predict the probable progression of examined data. High-quality and credible prediction can be derived only on the basis of a sufficiently large volume of data. Generally, the more relevant input data is available, the more accurate is the prediction of their progression. A common epiphenomenon of materials research, as well as the development of new products, is very limited amount of data. In such cases, the neural networks must be appropriately modified to achieve an acceptable accuracy of prediction based on small data sets.

Lolas and Olatunbosun (Lolas & Olatunbosum, 2008) used ANN to predict reliability behavior of an automotive vehicle at 6000 km based solely on information from testing the prototype. To this propose, they drawn up a three-phase optimization methodology for neural network development. The proposed network can detect degradation mechanism of the vehicle and use this knowledge to predict the trend of reliability throughout its life cycle. The overall error of the whole neural network and the three output parameters were less than 9%.

Li (Li & Yeh, 2008) deals with the prediction of a product life cycle already in the initial stages of manufacturing. For work with the small data sets, they developed nonparametric

learning algorithm named Trend and Potency Tracking Method (TPTM). This algorithm looks for the data trend by considering the occurrence order of the observed data and also quantifies the potency for each of the existing data by computing the TP value. It was experimentally verified that this algorithm helps to improve the performance of neural network prediction. This training mechanism conducts an incremental learning process and it is practical to earn the knowledge in dynamic early stages of manufacturing.
Li and Liu (Li & Liu, 2009) also deal with the use of neural networks for small data sets. They developed an unique neural network based on the concept of monitoring of a central data location (CLTM) for determining the network weights as the rules for learning. The experimental results confirmed the higher performance of prediction of the new network, especially in comparison with the traditional back-propagation neural networks.
Artificial neural networks offer very flexible technology broadly usable in the decision support systems. With the possibility of modification, neural networks are useful for applications which process a very limited amount of data. Laboratory research is one of the areas that are characterized by producing small data sets. One of the positive aspects of building a DSS for laboratory research using artificial neural networks is the fact that there are many commercial and free software tools for the design, development and implementation of the ANN.

3.5 Summary

The vast majority of above mentioned diagnostic DSSs are complex and robust tools with well-defined purpose, which processing huge amounts of data, representing hundreds or thousands of incidents and events, and having tens to hundreds of users. In contrast, DSS designed for use in research and development should be used by individuals with a maximum amount of data corresponding to tens of thousands events. The purpose of such system should be flexible in a certain manner, with the possibility of its definition according to the main objective of the research.
Artificial neural networks and Bayesian networks can be certainly considered as the appropriate technologies for development of DSS for laboratory research. Due to the relative inputs certainty of such system, the use of fuzzy logic seems to be somewhat excessive. Using of rule and case based reasoning logically seems to be inappropriate, mainly due to the absence or small number of already done cases, and the lack of rules for reasoning, especially in the early stages of research.

4. In-house DSS development overview

As was already mentioned in the introduction, this work is, among others, focused on in-house development as the best way to design, develop and implement DSS application with maximal possible compliance with user's demands and requirements. This part brings the short overview of in-house application development approaches and possibilities.
According to Turban (Turban et al., 2008), there are three basic approaches to DSS application development. They are:

1. build the system in-house,
2. buy an existing application,
3. lease software from application service provider (ASP).

Because of uniqueness of every research project, e.g. data types and data sources, amount of users, sharing and security requirements etc., there is basically only one solution to DSS

development – building the system in-house. In the case of laboratory research, buying an existing application brings requirements for adaptation to specific demands and requirements. In some cases, the effort to adapt the application could be comparable or even higher than the effort expended to the development on your own. Finally, lease application from the third party can provide satisfactory compliance with the requirements, but with higher financial costs.

There are two possibilities of building DSS application in-house – building from scratch, or building from components. Building from scratch is suitable for specialized applications. This option provides the best compliance with the user's requirements, but can be time-consuming and expensive. Building from components uses available, commercial, freeware or open source components, and creates the required application via their integration.

In-house application development includes several development approaches and techniques. The three most common are System Development Lifecycle (SDLC), Rapid Application Development (RAD) techniques, such as Prototyping, and End-User Development (Turban et. al., 2008).

SDLC is the traditional method of application development, mostly used for large DSS projects. It is a structured framework consisting of the follow-up processes by which an application is developed. Traditional SDLC consists of four basic phases – planning, analysis, design, and implementation (PADI), which lead to deployed system. Scheme of traditional SDLC is in Figure 4.

Rapid Application Development are methodologies for adapting SDLC, so the system can be develop quickly and some of the system functionalities can be available to users as soon as possible. It is an incremental development with permanent feedback from potential users. RAD methodology breaks a system into a number of versions that are developed sequentially. Each version has more features than the previous one, so the system is developed in steps.

The most widely used methodology of RAD is Prototyping. This methodology involves performing the stages of analysis, design, and implementation concurrently and repeatedly. After each increase in development, the system is presented to the potential users.

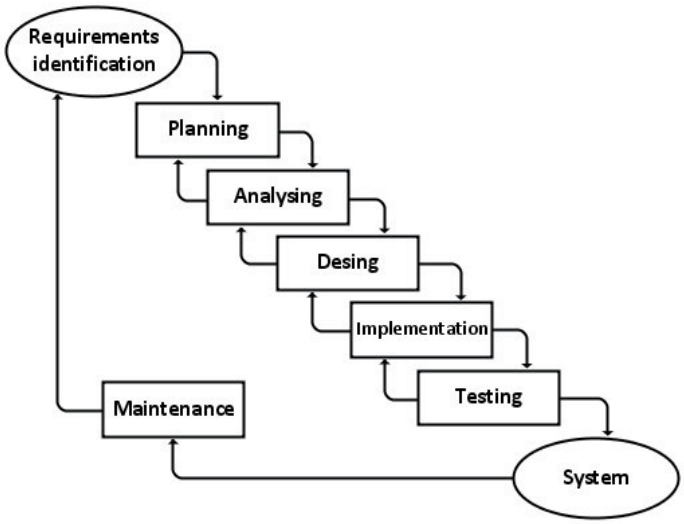


Fig. 4. Traditional application development approach - SDLC

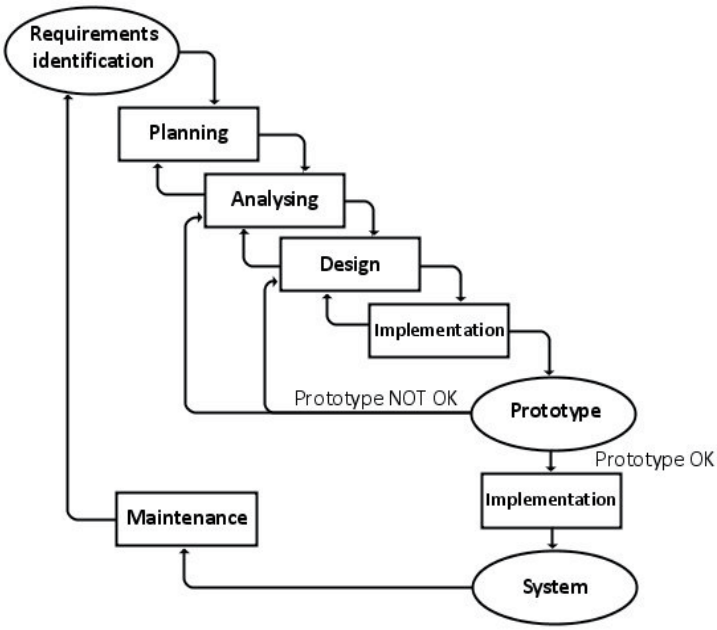


Fig. 5. Rapid Application Development methodology – Prototyping

Based on their response, further improvement takes place and the system is presented to the potential users again. After several iterations, no further improvements are proposed and the system is finally deployed. Scheme of Prototyping is in Figure 5.

End-User Development is a special case of DSS application development. The decision support tool is build by the users themselves. The advantages of this approach are undoubtedly the speed of development, the minimum cost and full compliance with the user's needs. On the other hand, there are considerable disadvantages associated with non-standard procedures of this type of development, such as inadequate documentation, improper use of development tools, the use of inappropriate technologies, or poor data security (Power, 2000).

5. Design and development of laboratory research DSS

The motivation for design and development of the decision support system for laboratory research is the need of practical tool for management and support of the research of organic semiconductors and their application in the field of vapor and gas sensors. The main aim of the proposed DSS is shortening the duration of research by revealing hidden knowledge in the measured data. Another goal is to make the research more efficient by the system's ability of manage, process, and properly present the measured data.

In the early stage, several system requirements were identified. The key requirements are: 1) requirement for presentation of measured data in a graphical form, 2) requirement for classification of the measured specimen, and 3) requirement for the prediction of material parameters of the sensitive layer according to the required response parameters.

There are also some requirements for the development. The first is to use in-house development. The second is to use only free or open source applications. Finally the third is

requirement for quick adjusting and implementing of the system functions. These requirements lead to selecting Rapid Prototyping as the development approach.

5.1 Design, architecture, and development possibilities

The common structure of DSS consists of four components – data management subsystem, model management subsystem, knowledge management subsystem, and user interface subsystem. The data management subsystem includes database and database management system (DBMS). It is usually connected to the data repository, such as data warehouse. The model management subsystem is software that includes quantitative models with analytical capabilities, and an appropriate software management. It can also include modeling languages for building custom models, and can be connected to the model storage. The knowledge management subsystem provides intelligence to facilitate the decision-making process. This subsystem is optional, but highly recommended. The user interface subsystem mediates communication between the users and the system. A Schematic view of the common DSS structure is shown in Figure 6.

According to requirements and based on the common structure of DSS, the architecture of proposed DS for laboratory research was designed. Architecture of proposed DSS consists of four interconnected components – database server, application server, web server, and graphical user interface (GUI). Some of these components include additional functional units. The architecture of proposed DSS is shown in Figure 7. Functions of single components and their roles are described below.

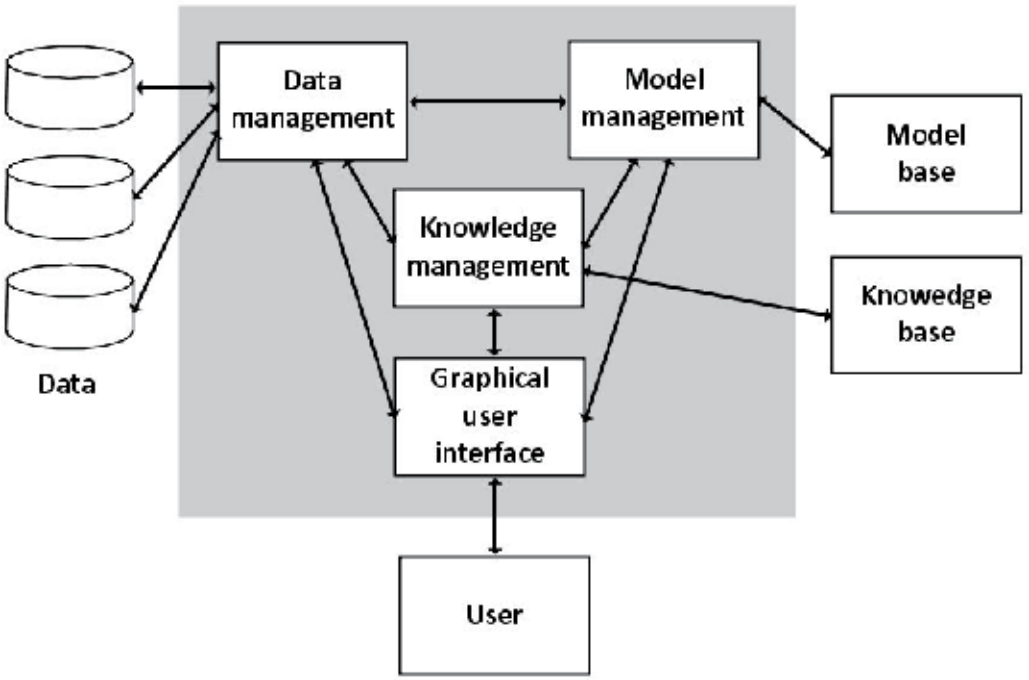


Fig. 6. A schematic view of DSS

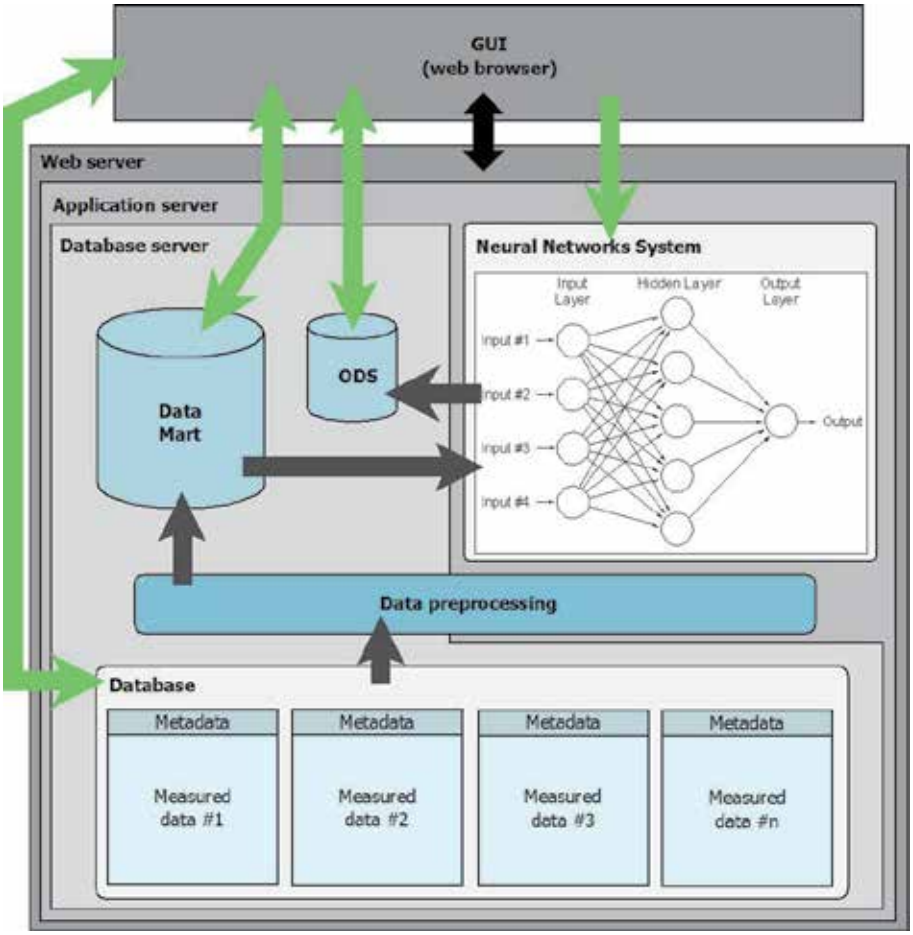


Fig. 7. Architecture of proposed compound DSS

5.1.1 Database server

The database server is a computer program that provides database services to other programs and computers. There are a number of free database servers. The most frequently used database servers are PostgreSQL, Sybase ASE Express Edition, MS SQL Server Express Edition, and Oracle Database 10g Express Edition, also known as Oracle XE. The basic criteria for choosing a database server are the dependence or independence on the platform (Windows, Unix, Mac OS), support of scripting and programming languages and standards, and the size and accessibility of the developer's community.

Based on these criteria, Oracle XE was selected as the database server for the proposed DSS. This solution provides a platform-independent, high quality and reliable database system, easy administration, and support for PHP, Java, .NET, XML and open source application. Oracle XE also involves PL/SQL language, a powerful tool for creating, storing and executing procedures applicable for the analysis and reporting. Due to the fact that Oracle is the database applications market leader, there is also large developer's community, providing support via the web discussion forums.

Within the database server, four functional units are built: relational database, data preprocessing unit, data mart (DMT) and operational data store (ODS). The main task of the relational database is to store data and metadata of all measurements and allows users to display information about the measurements and the measurements results in the form of graphs and charts. The data represent a set of measured values. The metadata contain information about particular measurement, such as the date and time of measurement, code of specimen, measurement conditions, used equipment, etc.

The data preprocessing unit, commonly known as ETL (Extract, Transform, and Load), is responsible for collecting and retrieving data from the relational database, their cleaning and adjusting into the desired form, and their subsequent loading into the data scheme of the data mart. The data have to be preprocessed for two reasons. First, the data have to be aggregated for further analysis, for instance using OLAP; second, the data have to be refined and consolidate for use in the neural network system.

Data preprocessing unit can be, with advantage, built using PL/SQL procedures. However, there are a lot of freeware ETL applications, such as Clove ETL, Pentago, Spago BI, KNIME, and more. In the case of using some of the open source applications, this functional unit will be move from the database server to the application server. It largely depends on the technology, respectively on the programming language of the application. This is also the cause of the unit overlapping between the database and application server in the diagram of the proposed DSS architecture, shown in Figure 7.

Data mart is a single subject data warehouse. Using a data warehouse technology allows performing quick ad hoc analysis, e.g. displaying data from different perspectives and in different contexts, such as used organic material, method of specimen preparation, measurement method, measurement conditions, etc. Operational data store (ODS) is the last functional unit of the database server. Its task is to store the data from the output of the neural network system and allows users to executing queries on these data.

5.1.2 Application server

The application server is a software framework designed for the effective implementation of procedures (programs, routines, scripts) to support building of applications. The task of application servers is to integrate heterogeneous environment when using multilayer architecture, and fully support the access to the various data sources. Major part of the application servers is based on the Java 2 Platform Enterprise Edition (J2EE) standard. The most popular open source application servers are Zope, JBoss, JOnAS, and GlassFish server OSE, which is a free version of the application server directly supported by Oracle.

Integration of the application server into the proposed DSS architecture mainly depends on the types and technologies of tools integrated in the neural networks system. The possibility of omitting the application server is strengthened by the fact that today's web and database servers taking over the function of application servers and allow direct integration of applications.

The main function of neural networks system is generation, preparation and application of neural networks for classification of measured data and prediction of their parameters and trends. For this purpose, it is necessary to choose a tool or tools that contain appropriate types of artificial neural networks, and are open-source enough to allow the modification of learning algorithms. There are many open source tools for creating neural networks, such as Fanny NuClass7, Joon, Encog, Neuroph, NNDef and more. Many of these are in the form of a library of some programming languages, like C + +, Perl, Python, .NET, or PHP.

5.1.3 Web server

Web server is the computer system responsible for processing requests from the client (Web browser) and transmitting the data in a network environment (Intranet, Internet), typically using HTTP (HyperText Transfer Protocol). Processing request means sending the web pages in the form of an HTML (HyperText Markup Language) document. Transmitted data can be static (prearranged data files, so called static content) or dynamic (dynamic content). Dynamic content is created on the client's request at the server side using different technologies (Perl, PHP, ASP, ASP.NET, JSP, etc.).

The proposed DSS will solely use the dynamic content. Technology used for creating the dynamic content will depend on the functions of the selected web server. Without a doubt, the most common open source web server is Apache. Other commonly used web servers are Roxen, Savant Server and nginx. There are also a number of web servers based on Java[2].

5.1.4 Graphical user interface

GUI is a very important part of the application. The quality of the user interface, its simplicity and clarity, significantly affects user's acceptation or rejection of the application. For creation of a graphical interface of proposed DSS, dynamic web pages displayed using web browser will be used. This solution is a logical consequence of the requirement for accessibility of DSS via intranet or internet.

Good possibility for building the whole DSS application is using of Integrated Development Environment (IDE), such as Eclipse or NetBeans. Both of these IDEs are Java based and support multiple programming languages. Eclipse can be used to develop applications in Java, C/C++, COBOL, Perl, PHP, Python, Ruby, Scala, Clojure, and Scheme. NetBeans includes all the tools needed for creation of professional desktop, enterprise, web, and mobile applications with the Java platform, as well as C/C++, PHP, Python, JavaScript, Groovy, Ruby, and others.

5.2 An overview of DSS for organic semiconductors research

The laboratory research of organic semiconductors, as well as any other research, produces relatively large amount of data in the form of measured values, information about measurement's properties and conditions and information about observed subjects. On the other hand, it also produces knowledge, either explicit or directly visible, or hidden in the data. To make the research efficient and effective, the data and knowledge have to be managed and used in an appropriate form. These reasons lead to development of a web-enable compound DSS which combines elements and characteristics of data-driven and knowledge-driven decision support system.

The objectives for development of the decision support system for laboratory research were build the system in-house and use only free or open source applications and tools. That leads to investigation and testing of several applications which were supposed to be suitable for DSS development. The most of them are already mentioned in previous sections. An overview of applications, tools, and techniques used for development of a compound DSS for organic semiconductors research is given below.

---

[2] For more details see http://Java-source.net/open-source/web-servers

The data-driven part of the DSS is based on Oracle XE and uses PL/SQL for data processing routines. This part is responsible for storing the raw measured data and their metadata as well as their preprocessing such as filtering, computing, transforming etc. It is also responsible for executing user's queries and analysis.

Due to relatively expensive and time-consuming measurements associated with organic semiconductors research, the main attention was paid to the knowledge-driven part of the DSS. The main aim of this part is satisfactory prediction of observed parameters based on several initial measurements. The secondary aim is discovering of hidden relationships and patterns in measured data.

For these purposes, the artificial neural network approach was chosen. The neural network system of the DSS is build using Neuroph – a Java neural network framework. This open source product allows development of common neural network architectures. Neuroph also provides GUI neural network editor and includes own IDE based on NetBeans platform.

For prediction tasks of the DSS, mainly the multilayer Perceptron (MLP) architecture with backpropagation training algorithms is used. The neural networks are deployed as Java applications. These ANN applications are integrated and executable using GlassFish Server OSE. It is an open source Java EE compatible application server. It is the free version of Oracle GlassFish Server. It provides more or less the same functionality with a broad support and developer community.

Undoubtedly, the best approach for building decision support system for laboratory research is evolutionary prototyping. Laboratory research projects are not static and it is clear that the requirements for decision-making support will still appear and change during all phases of research. Evolutionary prototyping approach enables developers to flexibly respond to current needs and requirements of users. In this way, the new functions and functionalities of the system could be implemented on demand and the system itself could be constantly up to date and satisfactory.

6. Conclusion

The usage of decision support systems in the field of laboratory research is still relatively unexplored area. The main aims of deployment of DSS for research purposes are shorten the duration of research and make the research more efficient. These objectives can be successfully achieved using artificial neural networks. Using DSS also brings the advantages in managing and processing of related data.

Such a system need to be built in the shortest possible time, and precisely tailored to the user's requirements. For these reasons, the in-house application development using evolutionary prototyping has been chosen as the most satisfactory approach. The architecture of proposed DSS consists of four interconnected components – database server, application server, web server, and graphical user interface. The application server is more or less optional, dependent mainly on functions of database and web server, and on the requirements of the neural network system.

This work proposes the approach to building decision support system for laboratory research. Based on characteristics, properties, and demands of laboratory research, the appropriate DSS types are discussed. Selection of applicable technology is derived from the capabilities of the four main categories of diagnostic DSS, used mainly in clinical medicine.

7. References

Aburas, H. M.; Cetiner, B. G. & Sari, M. (2010). Dengue confirmed-cases prediction: A neural network model. *Expert Systems with Applications*, Vol. 37, No. 6, (June 2010), pp. 4256-4260, ISSN 0957-4174

Cios, K. J.; Pedrycz, W.; Swiniarsky, R. W. & Kurgan, L. A. (2007). *Data Mining, a knowledge discovery approach*, Springer Science + Business Media, LLC, ISBN 978-0-387-33333-5

Faisal, T.; Ibrahim, F. & Taib, M. N. (2010). A noninvasive intelligent approach for predicting the risk in dengue patients. *Expert Systems with Applications*, Vol. 37, No. 3, (15 March 2010), pp. 2175-2181, ISSN 0957-4174

French, S.; Maule, J. & Papamichail, N. (2009). *Decision behaviour, analysis and support*, Cambridge University Press, ISBN 978-0-521-70978-1, Cambridge, UK

Gil, D.; Johnsson, M.; Chamizo, J. M. G.; Paya, A. S. & Fernandez, D. R. (2009). Application of artificial neural networks in the diagnosis of urological dysfunctions. *Expert Systems with Applications*, Vol. 36, No. 3, Part 2, (April 2009), pp. 5754-5760, ISSN 0957-4174

Kannappan, A.; Tilarasi, A. & Papageorgiou, E. I. (2010). Analyzing the performance of fuzzy cognitive maps with non-linear hebbian learning algorithm in predicting autistic disorder. *Expert Systems with Applications*, Vol. 38, No. 3, (March 2011), pp. 1282-1292, ISSN 0957-4174

Kumar, K. A.; Singh, Y. & Sanyal, S. (2009) Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU. *Expert Systems with Applications*, Vol. 36, No. 1, (January 2009), pp. 65-71, ISSN 0957-4174

Li, D. & Liu, C. (2009). A neural network weight determination model designed uniquely for small data set learning. *Expert Systems with Applications*, Vol. 36, No. 6, (August 2009), pp. 9853-9858, ISSN 0957-4174

Li, D. & Yeh, C., (2008). A non-parametric learning algorithm for small manufacturing data sets. *Expert Systems with Applications*, Vol. 34, No. 1, (January 2008), pp. 391-398, ISSN 0957-4174

Lindgaard, G.; Pyper, C.; Frize, M. & Walker, R. (2009). Does Bayes have it? Decision Support Systems in diagnostic medicine. *International Journal of Industrial Ergonomics*, Vol. 39, No. 3, (May 2009), pp. 524-532, ISSN 0169-8141

Liu, Y. & Li, S. (2007). Decision Support for Maintenance Management Using Bayesian Networks. *Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing*, ISBN 978-1-4244-1311-9, Shanghai, September 2007

Lin, C.; Lin, C. M.; Lin, B. & Yang, M. (2008). A decision support system for improving doctors' prescribing behavior. *Expert Systems with Applications*, Vol. 36, No. 4, (May 2009), pp. 7975-7984, ISSN 0957-4174

Lolas, S. & Olatunbosun, O. A. (2008). Prediction of vehicle reliability performance using artificial neural networks. *Expert Systems with Applications*, Vol. 34, No. 4, (May 2008), pp. 2360-2369, ISSN 0957-4174

NNMI lab. (2007). Bayesian networks, In: *NNMI Laboratory research*, 2010, Available from: < http://nnmi.kaist.ac.kr/res_contents/bayesian_eng.html>

Power, D. J. (2000). *Decision Support Systems Hyperbook*, Cedar Falls, IA: DSSResources.COM, Retrieved from <http://dssresources.com/dssbook/>

Ting, S. L.; Wang, W. M.; Kwok, S. K.; Tsang, A. H. C. & Lee, W. B. (2010). RACER: Rule-Associated Case-based Reasoning for supporting General Practitioners in prescription making. *Expert Systems with Applications*, Vol. 37, No. 12, (December 2010), pp. 8079-8089, ISSN 0957-4174

Tupa, J. (2008). *Control of diagnostics processes* (habilitation thesis), University of West Bohemia in Pilsen, Czech Republic

Turban, E.; Aronson, J. E.; Liang, T. & Sharda, R. (2008). *Decision Support and Business Intelligence Systems* (eight edition), Pearson Education, ISBN 0-13-158017-5, Upper Saddle River, New Jersey

Vikram, K. & Karjodkar, F. R. (2009). Decision Support Systems in Dental Decision Making: An Introduction. *Journal of Evidence Based Dental Practice*, Vol. 9, No. 2, (June 2009), pp. 73-76, ISSN 1532-3382

Zhuang, Z. Y.; Churilov, L.; Burstein, F. & Sikaris, K. (2009). Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *European Journal of Operational Research*, Vol. 195, No. 3, (16 June 2009), pp. 662-675, ISSN 0377-2217

25

Decision Support System for Designing with Polymer Materials – Current Challenges and Future Expectations

Urška Sancin and Bojan Dolšak
University of Maribor /
Faculty of Mechanical Engineering
Slovenia

1. Introduction

Product development process is rather precisely defined and reasonably well supported with modern computer tools. Numerous commercial computer aids or specially developed computer tools are relying mostly on graphic presentation, simulations, engineering analysis and animations of future product performance in virtual environment. In the embodiment phase, the designer has to take decisions, influenced by various parameters, according to the available data. One of crucial decisions is material selection, conditioned by several criteria, among which the focus is on function, technical features and shape of developing product. Other criteria, like serviceability, technical feasibility and economic justification are considered accordingly. Despite the potential of already mentioned computer tools, a designer has to evaluate the information gathered from these aids, seek interdependences and finally choose the optimum from the broad list of materials. Wide spectrum of various polymers or plastics should be outlined here, as they are frequently used in technical praxis. On the other hand, there is a leak of relevant data and knowledge for successful selection among them. Consequentially, the designer has to master all influential parameters, their overlapping or contradictions, and above all, he or she has to be acknowledged with materials available on the market. Designer's right/wrong decision severely influences product's applicability in praxis, its technical feasibility, life time, economic justification and recycling possibilities along with environment impact. Thus, numerous experts from various fields usually need to contribute their expertise to reach the final decision, which is often quite difficult, as opinions may be contradictive. In Small and Medium-sized Enterprises (SMEs'), the financial status commonly thwart the enterprise to hire an expert or entire team, which frequently leads to unprofessional design decisions. Young and inexperienced designers are also in arduous position at polymer material selection process. Solution to this quandary in product development process is the decision support system model for polymer material selection discussed in this chapter.

2. Product development process

The global market has adopted a continuous need to develop new, cost-effective, high-quality products at a very rapid and accelerating pace. To be able to compete, maximal

quality at minimal cost should be enterprises' motto. It has been assessed that 85% of problems with new products not working as they should, taking too long to bring to the market, or costing too much, are the result of a poor design process (Ullman, 2003). Thus, designers are often under pressure as they have to justify management's trust in new product also supported with diverse tools for selecting and evaluating the projects (Palcic & Lalic, 2009).

Product development is complex engineering process. The designer is progressing through the process dealing with many design and manufacturing problems, while envisage the production process (Vidal et al., 2005), product assembly, parts' maintenance, their influence on the environment or some other design aspects. Nevertheless, the designer's focus is on product's function where its form, materials and manufacturing processes are of equal importance beside it. In addition, the engineer has to define the new product features like tolerances, type of surface or material, from which the product will be produced. To solve these dilemmas, he or she has to rely upon own knowledge or experiences as existing Computer Aided Design (CAD) application does not provide any recommendations or guidelines.

Current CAD offers support during a great deal of engineering steps when designing a new product. CAD applications correspond to the designer's work at drafting, drawing, modelling, assembling, analysing and simulating. Its limitations appear, when having to accept certain determinations and decisions about the product. This is very important as the following steps of the process are directly or indirectly dependent on these decisions. In other words, the process is a sequence of interdependent events and one decision at the early design stage would then exert influence on all successive events, and the final design solution.

2.1 Decision-making in product development process

A product development process is above all a decision making process. The engineer has to choose the proper tools when performing the design process, such as selecting the adequate software for the initial problem and, more importantly, he or she has to make several decisions whilst working with these tools, in order to achieve an optimal solution. Human cognition plays the key role in product development, as knowledge domain is decisive during decision making process.

Experiences are an engineer's main advantage. Designers are in uncomfortable position here as at the beginning of their careers their experiences are limited. The possibility of acquiring experts' opinions is desirable, since they possess knowledge of specific design aspects and could contribute to the evaluation of possible design solutions. Moreover SMEs' often have quite such absence of knowledge due to economic capability required for hiring the specialists. This observation leads to the conclusion that adequate computer support is often needed for offering some advice and guidelines to designers during product development process. Such computer support for the decision process can be provided by the intelligent decision support system, presented in this chapter.

2.2 "Design for X" methodology

Within product development process, designers face many dilemmas linked with various aspects of the product, material selection being among them. It is one of its crucial decisions and is associated with many design and manufacturing problems, usually affected by basic

demands like application mode, and additional factors like supplier recommendation. The designer has to envisage the production process, semi-product or product assembly, parts' maintenance and evaluating the level of environmental influence, which is becoming increasingly important due to global pollution. Thus, Design for X (DfX) methodology, where X resembles the appropriateness for manufacturing, maintenance, service, etc. (Huang, 1996), has to be considered. The field of plastic products' design is of special interest regarding DfX, as the final product could be optimized for one or more domains of its life-cycle (Kuo et al., 2001). Using DfX, the engineer is able to concentrate on the most important domain within the life-cycle in order to provide an adequate product design solution.

Usage of DfX is rational and rationale as it represents one or the first step in the process of attaining optimal product design. Nevertheless, the designer still cannot expect any adequate support from available computer tools in the form of recommendation or guidelines when a material or technological procedure has to be selected. In order to overcome this bottleneck, Knowledge Base Engineering (KBE) techniques have to be taken into consideration, when developing an intelligent decision support system for polymer products' design. The intelligent system model for polymer products' design represents a new approach in design processing. The preliminary condition for knowledge-based support to polymer products' design process is an adequate knowledge base containing related, well organized DfX domain knowledge and relations.

3. Polymer materials

Conventional materials like metals and ceramics can be often substituted by others, more suitable for certain types of product. Thus, polymers are reasonable alternative as they could offer better characteristics for noticeable lower costs. Some of the advantages such as less weight, lower material costs, cheaper mass production, recyclability, specific electrical, isolative or, corrosive, etc. features, easier part joining (Kim, 2004), higher aesthetic values (e.g. no dyeing is needed) or, easier production of precision products, are of major importance in product development process.

For every scientific discussion it is important to clarify the terminology. Polymer literally means many units so polymers are materials where units consist of chain like macromolecules and are joined together through chemical binding. Furthermore, plastics are materials usually composed of polymers refined with various additives like fillers, glass fibres, and pigments which aggrandize polymers properties (Askeland & Fulay, 2009). Despite of described distinction between polymers and plastics authors usually use both terms interchangeably thus we adopted this idea in this chapter. Common classification of polymers is in three major groups:

- Thermoplastics including commodity and engineering polymers
- Thermosets materials
- Elastomers divided on natural and synthetic

General structure of thermoplastics is flexible linear chains, which could be straight or branched and direct the thermoplastics ductile behaviour. Characteristically for thermoplastics is their ability to melt at heating therefore they are processed in final form by heating at certain temperatures. Consequentially, they could be easily reformed and recycled with the same operation. Thermoplastics are generally distributed into

commodity and engineering polymers. First are of light weight with low strength and stiffness as well as not usable for applications at high temperatures and corrosion-resistant. They can be shaped without difficulty in various forms and are among inexpensive polymers. For mechanical engineers the engineering thermoplastics are in focus as they can offer enhanced strength, which in some cases can be greater than the strength of steel. Moreover, their performance at high temperatures sometimes as high as 350˚C is better but engineering thermoplastics' disadvantage is the expensiveness. Thermoplastics overall are present in various engineering applications like wire insulation, bottles, pipes and valves, automobile roofs, carpet fibres, packaging, egg cartons and windshields to mention just a few.

Thermosets have rigid three-dimensional network with linear or branched chains and are commonly stronger although more brittle then thermoplastics and sometimes more than metal or ceramics. As thermoplastics are easily shaped when exposed to temperature high enough, the thermosets to the contrary do not soft under heating but start to decompose. Thus, the recycling is problematic though very important due to the environment pollution. Thermosets are often applied as adhesives, coatings fillers, foams and laminates but are also present in applications like cookware and electrical mouldings.

Group of plastics with enormous elastic deformation are called elastomers generally known as rubbers with structure like thermoplastics or cross-linked thermosets consisting of spring-like molecules, which could be stretched by applying the force. Thermoplastic elastomers are special group of polymers and should be mentioned here as their processing is like thermoplastics' but the behaviour is elastic like elastomers'. Elastomer applications are diverse from tires and golf balls to hoses and seals.

To ensure the proper application of plastics, the designer has to consider three factors that determine the appropriate final use: design, processing and material selection. According to Material Information Society (ASM International, 2003), material selection factor has a great impact on all design aspects. Therefore some major designer's concerns are introduced:

- Designing products that can be built as easily and economically as possible
- Ensuring product reliability
- Simplifying product maintenance and extending product life
- Ensuring timely delivery of materials and components.

For successful product development is crucial to choose appropriate material, process and design matched to the part performance requirements.

4. Approaches to material selection

In contemporary engineering, the material and production process to be selected evolve and change whilst designing simultaneously with form of the developing product. During the process the product evolves from preliminary ideas and sketches and is refined in details, form, material, and production techniques. Being aware of this fact, the designer usually assemble as many information as possible studying the similar devices from the competition in order to understand the product and to set the requirements about the material, manufacturing process, assembly and maintenance. According to Ullman those information has a great impact on the embodiment phase of design at following domains (Ullman, 2003):

- Quantity of the product to be manufactured influences the manufacturing process selection to great extent as for very small series some manufacturing processes like injection moulding are not cost-effective.

- Prior-use knowledge can lead to past and effective solutions and at the same time the better material or manufacturing process solutions could be omitted.
- Knowledge and experiences are the key factors to top-level product contrarily the lack of both could lead in other direction. The invitation to the expert at material or manufacturing process selection is a regular practice in SMEs' to disseminate company's knowledge and experiences.
- Availability of the material is a very important point at material selection as some characteristics of the design could not be implemented with some always available cheaper materials.

4.1 Material selection methods

Material selection is a significant stage of the design process and a complex task, whose execution varies from enterprise to enterprise in accordance with staff and the economic aptitude of the company. In general, material selection methods can be, according to Ashby (Ashby & Johnson, 2005), arranged in four different selection methods called Selection by Analysis, Selection by Synthesis, Selection by Similarity and Selection by Inspiration. All methods require input data in the form of design requirements specific for each method. Selection by Analysis is the most systematic and robust as input requirements are objectives, functions, and constraints, and furthermore, they are precisely defined and unambiguous. Its deficiency derives from this particular distinctiveness, which causes the method to fail in the case of imprecise inputs or imperfectly formulated rules. Previous experience and analogy are key factors in the Selection by Synthesis method, where design requirements appear in the form of intentions, features, and perceptions. This method is used, when knowledge of the solved cases can be exploited and transferred to other product with some features in common. Selection by Similarity is the selection method, where input is already known or potential material solution and its purpose is to find substitutive material for an existing product, often initiated by design requirement changes due to e.g. environment legislation. The less uniformed method is Selection by Inspiration, where input is pure curiosity and the designer's task is to examine and analyse other solutions for a specific feature, in a systematic way. This method is used when no scientific method is helpful. All material selection methods and their variations are implemented in numerous variations as engineering praxis.

4.2 List of must and want properties in engineering praxis

List of must and want properties method can be characterized mostly as Selection by Analysis however it could be also applied as Selection by Synthesis or Selection by Similarity. Usually this method of material selection involves making a list of properties that you must have for future application and the list of properties that are desired for this particular application. These must and want properties are then matched with the properties of available polymer materials on the market. In engineering praxis, four basic groups of material properties are reviewed:

- Physical (specific heat, coefficient of thermal expansion, thermal conductivity, heat distortion temperature, glass transition temperature)
- Chemical (composition, additives, fillers, crystallinity, environmental degradation, spatial configuration, molecular weight, flammability)

- Mechanical (tensile and compressive properties, heat distortion, pressure-velocity limit, toughness, stress rupture resistance, creep resistance)
- Dimensional considering manufacturing conditions (manufacturing tolerances, stability, available sizes, moldability, surface texture)

In order to illustrate the importance of polymer materials' idiosyncrasies (Budinski & Budinski, 2010), each group of followed properties should be described. Physical properties are material characteristics that pertain to the interaction of these materials with various forms of energy and human senses. Generally they could be measured without destroying the material. Density is a physical property determined with weighting or measuring the volume of the product. Physical properties like feel and colour are even easier to determine while they affect the customer as he or she only looks at it. Nevertheless, they are not marginal material properties and their importance rises in today's consumer oriented society. The designer has to acknowledge that plastic feels different from metal and yellow is happier colour in comparison to brown.

Chemical properties are related to the structure of polymer material, its formation from the elements of which the material is made, its reactivity with chemicals and environments. These properties cannot be visually inspected and are measurable in chemical laboratory.

Mechanical properties are the features of material, which are put on view when it is exposed to a force. They are related to the elastic or plastic behaviour of the polymer and they often require destruction for measurement. Term mechanical is used because they are usually used to indicate the suitability of the material for use in mechanical applications – parts that carry a load, absorb shock, resist wear, etc.

Dimensional properties include as well manufacturing considerations like manufacturing tolerances and moldability. This category concerns also the surface texture and its roughness, which is measurable and essential for many applications. Available size, shape, finish and tolerances of the product are also important polymer material selection factors.

4.3 Material selection program packages

Regardless of the material selection method used in design process the designers and experts all should choose eventually from the broad list of materials in catalogues from several material suppliers or they can use the web or computer programs for easier material selection. Plastic material selection web programs like CAMPUS offer some comparison of plastic materials' properties for one or more suppliers. User can observe the relations between several property types like rheological, mechanical, thermal, processing, etc., where each type has some specific single value technical parameters introduced along with multi-joint graphical data at disposal. Although, the value of such computer tool for the engineer and design process is unequivocal, the plastic material selection support is limited. Firstly, the potential selected polymers – candidates are ranked according to some of basic properties, where maximum and minimum values are defined by the user. Depending on designer's skills to do so, the system offers from zero to hundreds of candidates, which meet the requirements. If the number of candidates is large, he or she has to define the values tightly and vice-versa, if there are no suggestions. Due to unambiguous ranking the user receives systematically acquired candidates. Thus, it could be misleading for the inexperienced designer as he or she could overlook the crucial parameters like time and temperature dependence of properties. In addition,

CAMPUS does not offer the designer the prices of the candidates and as we will be able to notice in this chapter the material cost is important and is checked numerous times during the polymer selection and design process. Another way of finding the most appropriate material is to investigate the polar charts, which allow the comparison of polymers for multiple selection criteria (Elsevier ed., 2010). The properties of each candidate are designated on n-scales, which radiate from a central origin yet the points on scales are joined to form the closed polygon. It is welcomed to define the minimum properties' values of the future product and to create its polygon so the possible candidates are only those, which enclose the product's polygon.

Sometimes during product design the requirements does not meet with existing materials or material production processes irrespective of designer's knowledge, experiences and effort. Consequentially, the project is postponed until new special designed material is developed to satisfy the design. It has to be subjected that this involves additional time and finances to support the project and that it is usually only justifiable for special products at high technology performance.

5. Plastic product design process

Design process is not always alike in all enterprises however the basic phases stay the same. While progressing through the task clarification, conceptual, embodiment and detail design phase engineer has to make numerous decisions not only about product design but also about manufacturing, assembly, maintenance, and its impact to the environment. Plastic product design is very distinguished (Alber et al., 2007) from conventional design with common materials like metal and ceramics as the plastics could offer better characteristics for noticeable lower costs. Some of the advantages mentioned before in Section 3 are of major importance. Therefore, material selection is one of key decisions, especially when the designer has to choose between approx. 120.000 different plastic materials (Ashby, 2005; Ashby & Johnson, 2005). It is of great importance as it influences technical and economical aspect of the product. To support this idea, a diagram of new product development process is described and illustrated in Fig.1. It also indicates where in the described development process, the previously selected material could be modified.

After 3D computer model of potential product is finished, a preliminary plastic material selection is set considering technical parameters, product type (product's purpose, exposure to the high temperatures, etc.), customer's requirements and wishes, fashion trend, special restrictions (product's contact with food, toys, etc.) and material costs. At this stage, first strain/stress analyses and casing simulations could be performed. Some severe material changes could be done according to results and usage of reinforcement fibres is discussed. Production process selection is next engineer's decision directly connected to plastic material selection as he or she has to consider products type, its size and precision, wall thickness, tolerances, surface roughness and even a size of production series. In case the manufacturer wants to use particular production machines or the series is large and the chosen polymer could be produced only with the production process suitable for small series, the need for material change is present again. Next phase is tool design, followed by detail casting simulation. Due to the results, some parameters like strain, elasticity, fibre reinforcement or thermal resistivity may have to be adjusted. Furthermore, additional analyses and simulations are fired to approve or to disprove the last selected plastic. Final step before sealed approval is calculation and costs appraisement.

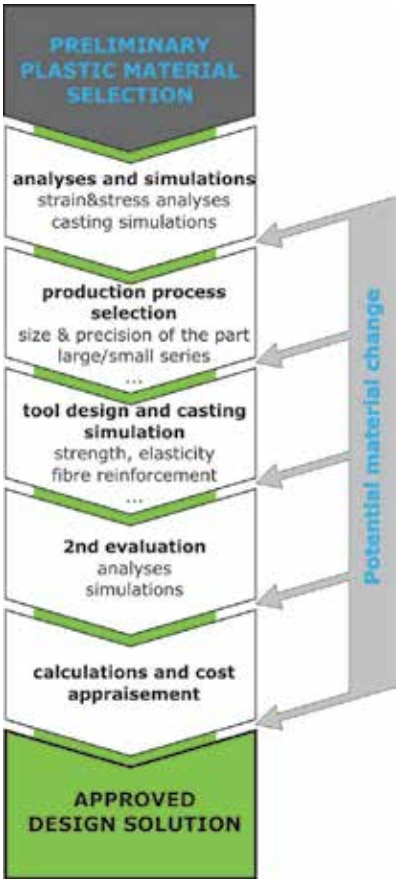


Fig. 1. New product development process

Evidentially from the Fig.1, the plastic material could be modified up to five times after preliminary polymer selection. In every phase, the designer has to make the best possible decision considering all relevant parameters, sometimes also contradictive. In order to surmount this obstacle and to make product development process less experience dependent, the decision support system for plastic products' design is proposed.

5.1 Material characteristics comparison – case-study

In general, there are diverse plastic materials on the market with similar technical features as widely used Actrylometril-butadiene-styrene (ABS). Designers are often used to design with it and do not consider any other material as potential material choice. High density Polyethylene (PE), Polypropylene, and Polyamide (also called Nylon) are some of the polymers, accurately thermoplastic polymers resembling overall characteristics but to a larger extent at certain parameters (Ashby & Johnson, 2005). Let's study one of the key parameters, the fracture toughness of these three materials in comparison to ABS (Fig.2). We can observe from the chart that Polyamide has the greatest extent of fracture toughness values, while Polypropylene has the smallest. Some versions of Polyamide and high density Polyethylene have lower fracture toughness than ABS but, simultaneously, these also have varieties with higher values of fracture toughness in regard to ABS. Moreover,

Polypropylene can also be tougher than ABS. To sum up, the toughest version of ABS still does not reach the value of the toughest versions of the other three materials (Sancin et al., 2010).

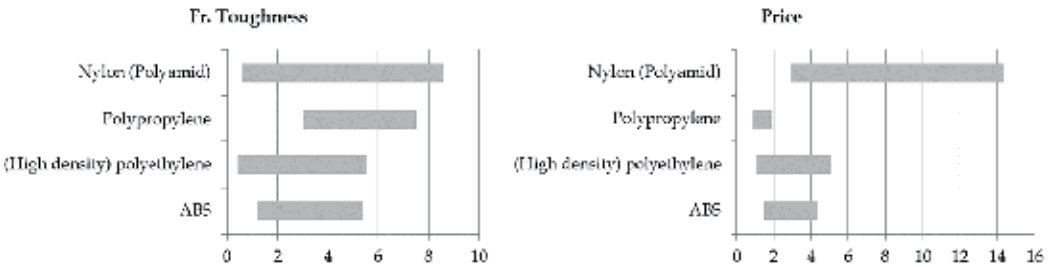


Fig. 2. Comparison of selected thermoplastics for fracture toughness and material price

In addition, the material price is an increasingly important parameter of the design process and the diagram on Fig.2 shows that Polypropylene and high density Polyethylene could have a lower price than ABS. It is necessary to explain that a higher price is sometimes justified due to the acquisition of other design components of the product, such as the possibility of smaller wall thickness, which leads to a reduction of the material needed for production of the manufactured goods. On the other hand, the designer has to evaluate numerous, and for the final product design crucial parameters within the development process by considering all the materials available on the market, with the aim of achieving optimal results, which can sometimes be an almost mission impossible. Younger inexperienced design engineers have major difficulties regarding material selection, thus, in order to overcome this barrier, the decision support advisory system for plastic product design is proposed. This computer aid will offer recommendations and guidelines according to the required parameters, shape or/and function of the product and could also be helpful for experienced designers using the system as a verification tool.

6. Intelligent decision support system for designing with plastics

Product design is a major task and a complex process. Its progress does not depend on just solving problems regarding known information like requested technical parameters, standards, conditions and other constraints that could be resolved with one ultimate solution but also on designers' experiences. Thus, the regular computer programs based on a database search do not fulfil the requirements of the designer the decision support system is a reasonable alternative.

Commonly, the decision support system model is executed in five phases, knowledge acquisition, decision support system shell selection, knowledge base construction, reasoning procedure definition, and development of a user interface. First and most time consuming phase is knowledge acquisition as experts' cognition in the field of design and plastics has to be preciously collected, adequately processed, well organized, and properly stored in the knowledge base. Before creating the knowledge base of the system containing human cognition, relations, and experiences, the decision support shell has to be selected. The generally accepted definition is that decision support systems are information systems supporting operational decision-making activities of a human decision maker. The decision support systems shall then help decision makers to compile useful information from raw

data that are distributed in a potentially heterogeneous IT infrastructure, personal or educational knowledge, models and strategies. As long as the system can be designed in advance and the focus lays on data, in practice the application of decision support systems is often sufficient. However, according to Yilmaz and Oren (Yilmaz & Oren eds., 2009) the versions of decision support systems are following:

- agent-directed decision support systems, used for rapidly changing environments where the system needs to be able to adapt;
- decision support simulation systems are used to obtain, display and evaluate operationally relevant data in agile contexts by executing models using operational data exploiting the full potential of modelling and simulation and producing numerical insight into the behaviour of complex system;
- agent-directed decision support simulation systems are agent-directed simulations that are applied as a decision support system, whose focus is on processes, so the system can adapt to new requirements and constraints in the environment.

For the decision support system for plastic material selection presented in Section 7 it is expected the agent-directed decision support simulation systems to cover all requirements and wishes the most satisfactory.

6.1 Current decision support systems for material selection

In recent years, many decision support systems where developed and some successfully launched in real applications (Turban et al., 2004). The significance of material selection dilemma within the design process is obvious, as several models were developed to support the engineers at this stage of design. It is essential to study and observe the progress on the field of material selection intelligent support thus, three examples of novel material selection models or methods, all for diverse applications, are going to be briefly discussed. It is significant to get the insight in the background before presentation of the future expectations in plastic material selection with decision support system.

Material selection query occurs at designing various products. Kumar and Singh represented intelligent system for selection of materials for progressive die components, where two knowledge base modules are comprised (Kumar & Singh, 2007). One is designated as selection of materials for active and inactive components of progressive dies. The other module was developed for determination of hardness range of materials for active dies. Acquired knowledge in the knowledge base is analysed and incorporated into a set of production IT rules.

At the present time, we acknowledge several formalised methods to support selection of individual materials. Real design problems often involve materials in combination as sometimes only joint multiple materials can produce the expected performance. Edwards and Dang address this issue and recommend a multiple-mapping strategy and an inter-level behavioural modelling strategy. Structural and materials solution in supporting design decision-making may be considered simultaneously (Edwards & Dang, 2007).

In general, two types of material selection methods were commonly used at past applications: material index based selection method and knowledge based methods. First follows some successive steps to identify the optimal material as the material heaving the largest/smallest material index value. When applying multiple indices the use of optimization method is mandatory to find the optimal result. Knowledge based methods use diverse approaches like IT rules approach, decision-making approach and fuzzy multi-

attribute decision-making approach. All demand a bunch of human cognition and intelligence just the last one capable of dealing with imprecise data significant for material selection. Ullah and Harib introduce novel material selection method, which does not require derivation of material indices or unpleasant inference calculations. Resembling the material index based selection method it uses always available material property charts as material relevant information in order to become realistic and user-friendly method. The application of presented method is selection of optimal materials for robotic components in early stage of design (Ullah & Harib, 2008).

6.1.1 Material selection results provided by decision support systems

As presented in Subsection 6.1, the decision support system for selection of material for progressive die components was developed alternatively to manual material selection with material and die handbooks, heuristics and designer's knowledge and experiences (Kumar & Singh, 2007). The adequate material choice is one of major activities in die components design leading to increased die life and consequential costs reduction of sheet material and costs of production. The proposed system differentiates from existing CAD systems for progressive dies at providing not only the list of materials but also an option to select easily available materials from the advice received from the system. After that the list of materials can be prepared appropriately. According to authors, the discussed system has been designated as powerful and easy to handle due to interactive mode of system-designer communication and extensive knowledge base containing knowledge and experiences of progressive die design. The system was developed to advice the designers at material selection for progressive die components, thus it is oriented mainly in just tool steels, which enables the flexibility of the system.

Edwards and Dang proposed a multiple-mapping strategy and an inter-level behavioural modelling strategy to support design decision-making when applying materials in combination and when they are combined with structural determination or design components (Edwards & Dang, 2007). According to the authors, the objective is to facilitate simultaneous consideration of components and materials at early stage of design and to provide the platform for the designer to work out the couplings among the material properties and corresponding components, in order to determine the respective material properties. Due to the complexity of research area discussed, proposed strategies cannot assure adequate support to the engineering design problems despite of their appropriate implementation and satisfactory validated case studies.

Intelligent materials selection method introduced by Ullah and Harib uses a linguistic description of material selection problems and material property charts relevant to the linguistic description of the problem (Ullah & Harib, 2008). The presented method supports the optimal material selection at early stage of design process and is suitable even for complex machineries, where design requirements and design relevant information are not precisely known. Discussed method could be applied for all groups of materials and is not limited to one.

It is obvious that material selection is interesting engineering domain and advanced decision making process. In recent years, numerous approaches and various methods were developed to create the decision support system for this particular field. Most material selection methods are specific at some points and offer a vast amount of opportunities to build diverse decision support systems. Some are limited to only one group of materials (Edwards & Dang, 2007; Ullah & Harib, 2008) and some are focused on special group of

engineering materials (Kumar & Singh, 2007). Decision support systems, which effectively support the designers at decision making process, are concentrated in knowledge base, specifically in human knowledge and experiences. Well organized and defined knowledge for material selection is of great importance and leads to more than adequate advice to designer and hence is a significant contribution to design process.

7. Intelligent decision support system for plastic product design – future expectation

The decision-making process is a constant for every designer aiming at a successful and efficient performance. Alternatively to experts' acquired domain knowledge, we decided to develop an intelligent decision support system (Đurić & Devedžić, 2002; Edwards & Deng, 2007; Kumar & Singh, 2007; Novak & Dolšak, 2008; Turban et al., 2004; Vitanov & Voutchkov, 2005; Zhu et al., 2008) in order to overcome the bottle neck - plastics material selection (Ullah & Harib, 2008). In the input data are a significant factor for intelligent module performance, the results of which depend on knowledge base content. The main objective of the proposed system is a consultancy with the designer in order to obtain the output, containing the most appropriate material for the product application, product design guidelines, etc.

The development methods included in research are a combination of human cognition in the field of design knowledge (Chen et al., 2007) and special domain knowledge expertise in the field of plastics. The knowledge base will contain human cognition useful for problem solving in the form of rules relating to modern plastic materials' selection and correlated manufacturing processes, assisted by the field of Design for Manufacturing (DfM) (Sevstjanov & Figat, 2007). Different approaches to knowledge acquisition (McMahon et al., 2004) and the appropriate formalisms for the presentation of acquired knowledge (Valls et al., 2009) within the computer program are of special importance. The potential for transparent and modular IF-THEN rules, whose advantage is neutral knowledge representation, uniform structure, separation of knowledge from its processing and possibility of dealing with incomplete and uncertain knowledge, is planned to be compared with more flexible knowledge presentation systems, such as fuzzy logic (Zio et al.,), where fuzzy sets and fuzzy rules will be defined as a part of an iterative process upgraded by evaluating and tuning the system to meet specified requirements. Tuning will be the most delicate job whilst building a fuzzy system as fuzzy sets and rules should frequently be adjusted during the system's construction. The main goal for the system is to apply domain knowledge, including human cognition, relations and experiences in the knowledge base of the system, which will, together with the data base, serviceable for a complex reasoning procedure (Benzmüller et al., 2008) behind the inference engine leading to qualified design recommendations and guidelines for designing plastic products.

The work on user interface development is in progress as research team devote a special attention to this issue, in order to enable transparent and efficient system application. Two different application modes have been anticipated, in regard to the type of input and output data. Guided mode (question and answer) will be used mostly at the beginning, when the first set of parameters has to be presented to the system. During the data processing phase, the system may present additional questions or ask for more parameters. In this case,

guided and graphic modes will be used to present the problem to the user. The solution in the final phase will also be presented in graphic mode.

7.1 Graphic environment of decision support system for designing with plastics

Characteristically for the designers is analytical thinking and when solving various engineering dilemmas they are used to study drawings, models, and simulations noticing all significant details. Virtual environment described in this section is one way of presenting plastic product design solutions to them in graphic mode as familiar working environment. Thus, it is expected graphic modes to be a part of future material selection decision support system applications.

Building the decision support system in graphic mode presented here aiming at a successful and efficient performance is a compound assignment. Research should be focused in human knowledge for problem solving formed as rules related to human cognition of modern plastic materials' selection and correlated manufacturing processes considering also DfM domain (Molcho et al., 2008). In addition, polymer material expertise together with human knowledge in the field of design process should fully be acquainted. In order to develop a graphic mode of plastic material selection decision support system, three major groups of polymer materials: thermoplastics, thermosets and elastomers, arranged in basic structure and presented in individual circles are proposed here (Fig.3). Within the framework of each circle several technical features carefully selected to cover all essential material properties, are assigned:

- Mechanical properties (strength, bending strength and working temperature),
- Production process (injection moulding, compression moulding, spin casting and extrusion),
- Chemical properties (resistance to base, acid, gas/oil, hot water),
- Working environment (internal/external use, fire resistance),
- Optical properties (colouring possibilities).

All three circles will have the same framework so the parameters, introduced to the system by the user will reflect through all of them. The system model will provide polymer material suggestions in discussed case, e.g. the designer receives four polymer material results, ABS, high density Polyethylene (PE), Polypropylene, and Polyamide, whose properties are introduced in the outer ring of the circle.

All design problems have a multitude of satisfactory solutions and almost never clear best solution. The significant feature of the system is two level results. The primary solution will be the possible plastic material choice, many of them or none. The database of the polymer materials will play the key role here. Afterwards, the system's knowledge base containing human cognition of plastic material selection and DfM will be of special importance as the system will be able to evaluate the candidates for potential material choices, which were just over the boundaries created by introduced parameters. Thus, some polymers are going to become a secondary solutions presented to the designer in form of notices containing recommendations about the advantages of each suggested solution. Considering the described decision support system with graphic mode, the enterprises will be able to compete at the global market by selecting the optimal material for their product. For the designers, especially for young and/or inexperienced ones, the plastic material selection decision support system is expected to be a helpful and indispensable tool as graphic mode

is anticipated to be a familiar and comfortable environment supporting engineering work on large scale in years to come.

Fig. 3. Example of virtual environment for plastic material selection

8. Conclusion

Modern engineering work is computer dependent. Although computer tools assist the engineer during product development there are still limitations at offering some advice or recommendation at decision-making process. Designer's aim is to have as many optimal solutions as possible, meaning results on diverse queries from designing a product shape, defining its tolerances or choosing adequate material. Plastics are extensively present in everyday life. However it is not a trivial decision when selecting polymer material for a new product. Due to wide spectrum of plastics at disposal, the plastic material selection is about knowledge and experiences, where inexperienced designers are in thwart position at decision-making. SMEs' as well are sometimes kept in the background in case they do not possess the polymer engineer and are forced in hiring an expert. This chapter is an attempt of presenting the engineering dilemma at plastic material selection. Several system models with different applications on material selection are also introduced to enlighten the state-of-the-art in the field discussed. Future expectations from our side at decision support system for plastic products are also explained. This idea is then upgraded with application of graphic mode, which is expected to be added value to the system as designers will have the opportunity to work in reliable, user-friendly environment. Consequentially, the design process is anticipated to be faster, more efficient, and less experience-dependent.

9. Acknowledgment

The special acknowledgement goes to students of industrial design in 4th year of study, who provide the basic idea of graphical appearance of the decision support system for plastic

products' design within the student project in educational process at the Faculty of Mechanical Engineering, University of Maribor.

10. References

Alber, B., Rieg, F. & Hackenschmidt, R. (2007). Product Design with High-tech-polymers – Practical Use of CAE-tools with Cross-linked Simulations and Experimental Verification. *Materialprüfung*, Vol.49, No. 7-8, pp. 402-407, ISSN 0025-5300

Ashby, M. F. (2005). *Materials selection in mechanical design* (3rd ed.), Elsevier, ISBN 0-7506-6168-2

Ashby, M.F., & Johnson, K. (2005). *Materials and design*. Elsevier, ISBN 0-7506-5554-2

Askeland, D.R. & Fulay P.P. (2009). *Essentials of Materials Science & Engineering* (2nd ed.), CL-Engineering, ISBN 0-4952-4446-5

ASM International (2003). Characterization and Failure Analysis of Plastics, ASM International , ISBN 0-87170-789-6

Benzmüller, C., Sorge, V., Jamnik, M., & Kerber, M. (2008). Combined Reasoning by Automated Cooperation. *Journal of Applied Logic*, Vol.6, pp. 318-342, ISSN 1570-8683

Budinski, G.K. & Budinski, M.K. (2010). *Engineering Materials* (9th ed.), Pearson Prentice Hall, ISBN 0-13-610950-0

Chen, R.Y., Sheu, D.D., Liu, C.M. (2007). Vague Knowledge Search in the Design for Outsourcing Using Fuzzy Decision Tree. *Computers & Operations Research*, Vol.34, pp. 3628-3637, ISSN 0305-0548

Đurić, M., & Devedžić, V. (2002). I-Promise—intelligent Protective Material Selection. *Expert Systems with Applications*, Vol.23, pp. 219-227, ISSN 0957-4174

Edwards, K.L., & Deng, Y.-M. (2007). Supporting Design Decision-making when Applying Materials in Combination. *Materials & Design*, Vol.28, pp. 1288-1297, ISSN 0261-3069

Elsevier ed. (2010). Design Engineering Manual, Elsevier, ISBN 978-1-85617-838-9

Huang, G.Q. (Ed). (1996). *Design for X - Concurrent Engineering Imperatives*. Chapman & Hall, ISBN 0-4127-8750-4

Kim, I.S., Prasad, Y.K.D.V., & Stoynov, L.A. (2004). A study on an Intelligent System to Predict the Tensile Stress in Welding Using Solar Energy Concentration. *Journal of Materials Processing Technology*, Vol.153-154, pp. 649-653, ISSN 0924-0136

Kuo, T.-C., Huang, S.H., & Zhang, H.-C. (2001). Design for Manufacture and Design for 'X': Concepts, Applications and Perspectives. *Computers & Industrial Engineering*, Vol.41, pp. 241-260, ISSN 0360-8352

Kumar, S., & Singh, R. (2007). A Short Note on an Intelligent System for Selection of Materials for Progressive Die Components. *Journal of Materials Processing Technology*, Vol.182, pp. 456-461, ISSN 0924-0136

McMahon, C., Lowe, A., & Culley, S. (2004). Knowledge Management in Engineering Design: Personalization and Codification. *Journal of Engineering Design*, Vol.15, No. 4, pp. 307-325, ISSN 0954-4828

Molcho, G., Zipori, Y., Schneor, R., Rosen, O., Goldstein, D., & Shpitalni, M. (2008). Computer Aided Manufacturability Analysis: Closing the Knowledge Gap between the Designer and the Manufacturer. *CIRP Annals - Manufacturing Technology*, Vol.57, pp. 153-158, ISSN 0007-8506

Novak, M., & Dolšak, B. (2008). Intelligent FEA-based Design Improvement. *Engineering Applications of Artificial Intelligence*, Vol.21, pp. 1239-1254, ISSN 0952-1976

Palcic, I., & Lalic, B. (2009). Analytical Hierarchy Process as a Tool for Selecting and Evaluating Projects. *International Journal of Simulation Modelling*, Vol.8, pp. 16-26, ISSN 1726-4529

Sancin, U., Dobravc, M. & Dolšak, B. (2010). Human Cognition as an Intelligent Decision Support System for Plastic Products' Design. *Expert Systems with Applications*, Vol.37, No.10, pp. 7227-7233, ISSN 0957-4174

Sevstjanov, P., & Figat, P. (2007). Aggregation of Aggregating Modes in MCDM: Synthesis of Type 2 and Level 2 fuzzy sets. *Omega*, Vol.35, pp. 505-523, ISSN 0305-0483

Turban, E., Aronson, J.E., & Liang, T.P. (2004). *Decision Support Systems and Intelligent Systems*, Prentice Hall, ISBN 0-1304-6106-7

Ullah, S.A.M.M., & Harib, K.H. (2008). An Intelligent Method for Selecting Optimal Materials and its Application. *Advanced Engineering Informatics*, Vol.22, pp. 473-483, ISSN 1474-0346

Ullman, D.G. (2003). The Mechanical Design Process (3rd ed.), McGraw-Hill, ISBN 0-0723-7338-5

Valls, A., Batet, M., & Lopez, E.M. (2009). Using Expert's Rules as Background Knowledge in the ClusDM Methodology. *European Journal of Operational Research*, Vol.195, pp. 864-875, ISSN 0377-2217

Vidal, A., Alberti, M., Ciurana, J., & Casadesus, M. (2005). A Decision Support System for Optimising the Selection of Parameters when Planning Milling Operations. *International Journal of Machine Tools & Manufacture*, Vol.45, pp. 201-210, ISSN 0890-6955

Vitanov, V.I., & Voutchkov, I.I. (2005). Process Parameters Selection for Friction Surfacing Applications Using Intelligent Decision Support. *Journal of Materials Processing Technology*, Vol.159, pp. 27-32, ISSN 0924-0136

Yilmaz, L. & Oren, T. (Eds.). (2009). *Agent-directed simulation and systems engineering*, John Wiley & Sons, ISBN 3-5274-0781-2

Zhu, F., Lu, G., & Zou, R. (2008). On the Development of a Knowledge-based Design Support System for Energy Absorbers. *Materials & Design*, Vol.29, pp. 484-491, ISSN 0261-3069

Zio, E., Baraldi, P., Librizzi, M., Podofillini, L., & Dang, V.N. (2009). A Fuzzy Set-based Approach for Modeling Dependence among Human Errors. *Fuzzy Sets and Systems*, Vol.160, pp. 1947-1964, ISSN 0165-0114

26

Decision Support System for Sustainability Assessment of Power Generation Technologies

Jūratė Šliogerienė[1], Dalia Štreimikienė[2] and Artūras Kaklauskas[1]
[1]*Vilnius Gediminas Technical University, LT-10223 Vilnius*
[2]*Lithuanian Energy Institute, LT-44403 Kaunas*
Lithuania

1. Introduction

The necessity to achieve sustainable economic development, which would be environment-friendly, would conserve natural resources and would not contribute to social tensions, is increasingly the key attitude in development strategies and plans for a range of economic activities, and in search for the best solutions. Sustainable development is understood as a lasting ideology of social change—as a compromise, which reconciles environmental, economic and social goals of our society. In the context of sustainable development, the energy development—the ability to ensure sufficient energy sources to the public—is of particular importance. Sustainable energy development is a complete set of measures, including better performance of energy production and consumption, gradually decreasing consumption of fossil fuels, lower pollution, introduction of renewable energy sources and advanced energy technologies, ensuring socially just pricing and accessible energy. The future of energy must go hand in hand with the concept of sustainable development and must ensure economic development of the public. Lately, many European countries face the issues of growing energy demand, and the consequences of global warming, ever-higher import dependency, also high and fluctuating prices of resources and energy. These issues lead to revisions of development projects and social programmes in the energy sector, and encourage adoption of instruments able to reduce social tensions, to satisfy the demand and to improve social safety.

The EU energy policy is a means to ensure secure, competitive and sustainable energy. The document *Green Paper: A European Strategy for sustainable, competitive and secure energy* published by the European Commission on 8 March 2006 sets forth the key areas with specific energy development challenges. This document prescribes the axes for energy development, and contribution to economic growth and job creation in Europe. It also names the challenges to secure safe energy supply in the internal market, as well as the solidarity of member states, promotion of competitiveness, increasingly efficient and diverse energy, and innovation (Commission of the European...2006). Building of relevant strategies and development plans, selection of specific instruments in line with environmental conditions are among the priorities in each state.

Development based on the sustainability principles, as well as planning, building and validation of various strategic decisions, demands for analysis and assessment of versatile information, such as EU policies and guidelines; political, social, economic and

environmental factors, and their changes; technical and technological data; and information from diverse stakeholder groups with different goals. When a decision-making process starts and we have to analyse diverse data, sometimes hardly fit for comparison, we want adequate tools, which help to consider our changing environment, to identify the weights of the defining criteria, and to reconcile economic, environmental, technological, social and other aspects. The information is abundant and often contradictory, thus only modern multiple criteria evaluation methods, based on the mathematical analysis of data, and integrated software applications make its assessment possible.

2. Multiple criteria analysis methods, their application

When it comes to handling of diverse practical tasks—notably to building of development scenarios, strategic scenarios and investment projects—one must consider economic, environmental, technological, social and other aspects, assess possible alternatives and rank them in the selected order of priority (Figueira *et al.* 2005). It may be selection of the best investment ortechnological project, analysis of alternative scenarios, environmental assessment of different regions in a quest for investment opportunities, etc. More and more books offer decision-making based on multiple criteria analysis methods. Effective decisions are of particular importance in investment projects aiming to ensure provision to the public, in projects financed by national authorities, and in cases that warrant objectivity, transparency and minimised influence of stakeholders.

A number of multiple criteria analysis and evaluation methods have been developed, and are used, worldwide. The newest multiple criteria evaluation methods, above all, facilitate assessment and comparison of objects described by both quantitative and qualitative criteria, by indicators expressed in different units of measurement (Scholz & Tietje, 2002). The scientists Gutsche, Zimmermann built and described typical systems of indicators, which can be used in multiple criteria analysis methods to measure the differences of quantity, quality and market conjuncture between a comparable object and a valuated object (Zimmermann & Gutsche, 1991). Ming-Te Lu developed an expert system, which uses a system of criteria and helps to select the most profitable or adequate real estate projects (Lu & So, 2005). Hwang and Quigley suggest a price assessment model based on the analysis of hedonic and sales comparison approaches; the model helps to estimate an efficient price indicator (Hwang & Quigley, 2004). Multiple criteria decision-making methods were used for evaluation of external services available to companies (Almeida, 2005), for evaluation of risk management and management of water supply systems (Morais & Almeida,. 2007), for project risk assessment (Zeng et al., 2007), for risk evaluation in natural gas supply systems (Brito & Almeida, 2008), and to reconcile infrastructure investments with environmental problems (Higgins et al., 2008). Multiple criteria methods are an attempt to choose an optimal decision when the alternative decisions must be concurrently assessed based on several contradictory criteria. Multiple criteria analysis methods view alternatives in an integrated manner: they deal with quantitative (operational territory, number of objects, cost, expenses, production capacity, etc.) and qualitative (legal acts, regulations, restrictions, stakeholder influence, technological novelty, compliance with environmental requirements, level of innovation, etc.) criteria of the current market conjuncture that describe the value of the item in question. Many current tasks related to energy strategy, development and technology selection are multiple criteria tasks.

Sources of literature, which discuss multiple criteria evaluation methods, also suggest conditional classifications: methods are classified according to the sets of alternatives, the

units of measurement, the decision-making rules, and the standardisation of evaluation results. For instance, by the type of data, decision-making methods may be classified into three groups: 1) deterministic, 2) stochastic, and 3) fuzzy sets. There may be cases, however, when different types of data are combined. Many authors suggest classifications which generally differ only by the comprehensiveness and number of methods. The key difference between classifications suggested by various authors is that some classify methods only by the type of information about indicators, while others introduce categories of information about alternatives (Chen & Hwang, 1991; Hwang & Yoon, 1981; Triantaphyllou, 2000).

Multiple criteria decision-making methods are most often classified into two distinct groups with different methodology for identification of preferences and for aggregation of information about criteria (Hwang & Yoon, 1981; Zavadskas et al., 1994). The first group includes multiple criteria methods from the value (utility) theory based on the premise of compensation — comparison of criteria: a possibility to fully balance the negative aspects of one criterion with positive aspects of another . The other group includes the outranking methods based on the concept of value without compensation and denies that criteria may offset one another. This methods may be further classified into three subgroups: 1) selection of the most beneficial variant using the utility function, 2) compromise models for selection of the variant closest to ideal, 3) concordance models to determine the priority relations of the highest compatibility (Hwang & Yoon, 1981; Zavadskas et al., 1994; Guitouni & Martel, 1998; Jeroen, 1999).

Multiple criteria analysis methods are abundant; their choice is based on the available data, goals, desired result and participation of decision-makers in the evaluation process. We shall proceed with a brief review of several multiple criteria methods, which are most adequate worldwide and are best suited for environmental analysis, for evaluation of project alternatives and technologies in the energy sector, and for integrated handling of environmental issues.

1. Multiple criteria methods of the value (utility) theory. This group of methods uses qualitative input and produces quantitative output. The group has two main subgroups: analytic hierarchy process methods and fuzzy set methods. The *Analytic Hierarchy Process (AHP)* was developed by the American scientist Thomas A. Saaty; lately, it is gaining popularity worldwide and is the most frequently used method for paired comparison of indicators (criteria, objects, features). It helps to find the weights of indicators located on the same level of a hierarchy with respect to a higher level or weights of hierarchically unstructured indicators. This method is based on a paired comparison matrix. Experts compare pairs of all indicators (technologies) in question R_i and R_j $(i, j = 1,...,m)$; here m is the number of compared indicators (features).

It is a convenient method, because paired comparison of indicators is simpler than comparison of all at once. The comparison of indicators is simple and rather reliable: it reveals the degree to which one indicator is more important than the other. This method enables transformation of a qualitative expert assessment of indicators into quantitative assessment. Such comparison produces a quantum matrix $P = \|p_{ij}\|$ $(i, j = 1,...,m)$. Mr Saaty suggests evaluations using a 5-point scale (1-3-5-7-9), which is frequently used in real-life applications (Saaty, 2000; Tam et al., 2006).

Multipurpose problems need to be separated into several components, because it helps to simplify the problem and to structure it better. A hierarchy with different goals and/or

layers of instruments must be compiled for each problem. This method is handy when one has to deal with problems hard to define and to assess expert opinions to be later used in problem solving. Moreover, the method is better at rendering the processes of human thinking than the method of logical strings. Besides being handy in finding the best solution, the Analytic Hierarchy Process also facilitates qualitative expression of priorities with the help of outranking tools.

The *Graphical Evaluation Method* is handy for visual representations of information related to facts identified after assessment of alternatives. Graphical visualisation of information also helps to determine the interrelations in and the structure of a phenomenon, and is useful in comparison of alternatives with several criteria, because it helps to visualise the interrelations between the respective criteria (Bertin, 1981; Khuri, 2002).

Sensitivity Analysis. Whereas the comparative scores and priorities of criteria are undefined in many comparisons of alternatives, evaluations and the selected valuation techniques are based on different premises. Since any evaluation aims to provide a decision-maker with the best alternative or a ranking of alternatives, such uncertainties are important only in assessment of their effect on the ranking. The decision-maker should find to what extent (percentage) the actual values could deviate from the values in the tables for effect evaluation or in the set of weights. The method offers the probability ranking of alternatives, which may be used in the analysis of ranking sensitivity of alternatives considering the overall uncertainty of the effects and priorities (Tam et al., 2006)

2. Outranking methods. Both input and output of these methods is quantitative. This group includes multiple criteria methods of the utility theory and a number of other types: TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*), SAW (*Simple additive Weighting*), LINMAP (*Linear Programming Techniques for Multidimensional Analysis of Preference*), ELECTRE (*Elimination Et Choix Traduisant la Realite*), PROMETHEE I, II, MELCHIOR, ORESTE, COPRAS, ARAS, etc. The methods in this group have a strict mathematical foundation on axioms. They are convenient because each alternative has its utility expressed in a quantitative form, and the comparison of values is simple. But these methods have a drawback: quantitative measurements are prone to inaccuracies due to slips by respondents or to other types of errors (Streimikiene & Mikalauskiene, 2009). When this group of methods is used, the results produced by various criteria are ranked and then the rankings are analysed. The outranking method is based on paired comparison of alternatives. All pairs for a criterion in question must be compared. The better alternative of each pair is determined by summing the results according to all criteria. This simple technique is used for quantitative data. Qualitative data, if any, are interpreted as unknown quantitative weights. The set S must be defined to include all strings of quantitative weights matching the qualitative priority information. Sometimes one alternative will be preferred from the entire set S, and in other cases one alternative may be preferred only from a certain part of the set S, with preference given to other alternatives in other parts of the said set. The distribution of weights in the set S is deemed unchangeable; the relative values of subsets in the set S may, therefore, be interpreted as a probability that one alternative in each pair is always preferable over the other. Probabilities are then summed to rank general alternatives (Von Winterfeldt & Edwards, 1986).

The *PROMETHEE* method differs from other multiple criteria methods with its deeper logics. The method is based on the so-called priority functions. Decision-makers may select these functions and set their parameters themselves. The PROMETHEE method offers a wide selection of functions to enable better reflection of the evaluator's opinions. At the

basis of the method, there is the matrix $R = ||r_{ij}||$ compiled from the defining indicators of compared objects and statistical data (or expert assessment), and weights of the indicators ω_i; $(i = 1, 2,..., m;\ j = 1, 2, ...,n,$ here m is the number of indicators and n is the number of compared objects, i.e. alternatives).

Quantitative multiple criteria evaluation methods determine whether an indicator is maximising or minimising. The best values of maximising indicators are the highest, while the best values of minimising indicators are the lowest. The criteria in quantitative multiple criteria methods often combine normalised values and weights of indicators. The logics employed in the PROMETHEE method differs from other quantitative multiple criteria methods. The decision-maker is an active participant in the phase of problem shaping and problem solving. The decision-maker adds the priorities in the method's assessment procedure: determines the permissible extremes of differences q and s (highest and lowest) for each indicator (criterion) R_i. In the PROMETHEE method, alternatives A_j and A_k are considered indifferent with respect to the indicator R_i, if the difference $d_i\ (A_j,A_k) = r_{ij} - r_{ik}$ between the indicator's values r_{ij} and r_{ik} is below the lowest extreme value q. Also, the alternative A_j is preferred over the alternative A_k if the difference is above the highest extreme value s. Moreover, Decision maker sets a specific priority function $p(d)$ (with the parameters q and s) for each indicator. The function's values vary between 0 and 1, and show the extent to which the alternative A_j is more important than the alternative A_k (with respect to the indicator R_i). In practical applications, six variants of typical priority functions $p(d)$suffice (Podvezko & Podvezko, 2009). The PROMETHEE method bases its final evaluation on all positive priorities of each alternative. The PROMETHEE I method defines the relation of priority and indifference for all alternatives A_j and A_k with either plus or minus: P^+, P^-, I^+, I^-. The PROMETHEE II method ranks the alternatives by the differences $F_j = F^+_j - F^-_j$. The PROMETHEE I method determines the best of compared alternatives (Brans & Mareschal, 2005).

COmplex Proportional ASsessment method (*COPRAS*). This method was developed by K. Zavadskas, F. and A. Kaklauskas (1994), scientists from Vilnius Gediminas Technical University. In multiple criteria analysis, it is expedient to combine the quantitative and qualitative assessment. It is the COPRAS method that helps to analyse more aspects of one object by combining quantitative and qualitative criteria. This method has a huge advantage—it helps not only to compare any alternatives but also to measure their market value. In this method, the selected alternatives are subjected to integrated analysis, considering quantitative (e.g., operational territory, number of objects, cost of products or services, production capacity, the replacement cost) and qualitative (e.g., restrictions imposed by legislation and regulations, technological novelty, compliance with environmental requirements, level of innovation, stakeholder influence, etc.) criteria of the market conjuncture, which describe the object.

It is easy to express quantitative criteria by the quantitative measures of your choice (amount of money, technical parameters, etc.), but in case of qualitative criteria expressed by conditional measures (scores, percentages) it is a more complex procedure to measure their values and weights. Weights for qualitative criteria must be identified through analysis, scientific studies and databases, by comparing equivalents, and by analysing macro-, meso- and microenvironment in regions with similar development degrees or development trends. Comparability demands for normalisation of the values of quantitative and qualitative criteria using relevant formulas. The expert assessment method is the most popular when it

comes to identification of weights for criteria from a variety of fields. When expert methods do their job and we have the criteria weights, we may identify their priority in the order of importance. Although expert methods do not secure very accurate values of quantitative criteria, the integrated method for calculation of criteria weights considers their qualitative and quantitative characteristics. Weights may be assessed using the outranking method. The COPRAS method proceeds with calculations in the following sequence (Fig. 1):

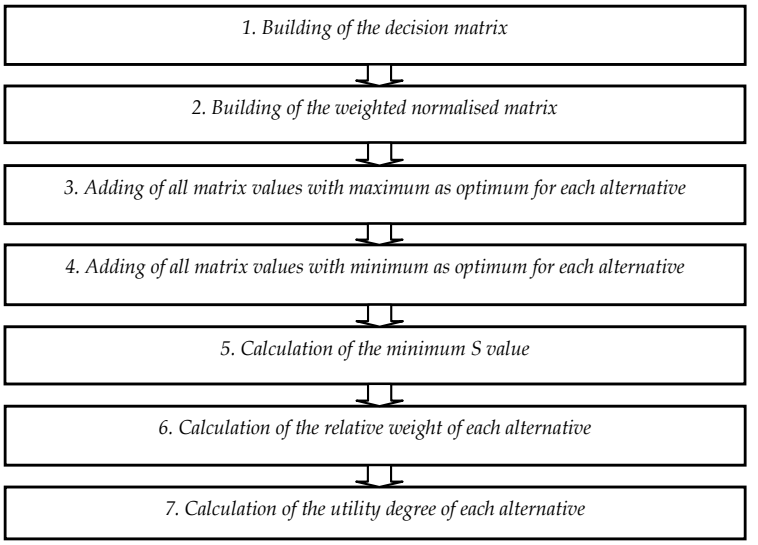


Fig. 1. The algorithm of calculations in the COPRAS method

Problem framing is the same as in other multiple criteria evaluation methods; it starts with building an initial decision-making matrix from n alternatives described by m indicators, thus $\{fij\}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. The weights of the indicators will be qi, i = 1, …, m.

Weighted normalized matrix. Comparison of indicators expressed in different measures demands for transformation of the indicators into dimensionless (normalised) values. Then a weighted normalised decision-making matrix is compiled. This stage aims to obtain dimensionless (normalised) weighted values from the comparative indicators. When the dimensionless weighted values are known, all indicators expressed in different measures may be compared.

The following formula is used for this purpose:

$$d_{ij} = \frac{x_{ij} \cdot q_i}{\sum_{j=1}^{n} x_{ij}}, \quad i = \overline{1,m}; \; j = \overline{1,n}, \tag{1}$$

here x_{ij} is the value of the criterion i in the alternative solution j; m is the number of criteria; n is the number of alternatives; q_i is the weight of the criterion i. The sum total of dimensionless weighted values d_{ij} for each criterion x_i is always equal to the weight q_i of this criterion:

$$q_i = \sum_{j=1}^{n} d_{ij}, \quad i = \overline{1,m}; \quad j = \overline{1,n}. \tag{2}$$

The value of the weight q_i of the criterion in question is proportionally distributed to all alternatives according to their values x_{ij}. Now the sums of weighted normalized minimising S_{-j} and maximising S_{+j} indicators describing the alternatives are calculated. The following formulas are used:

$$S_{+j} = \sum_{i=1}^{m} d_{+ij}; \tag{3}$$

$$S_{-j} = \sum_{i=1}^{m} d_{-ij}; \quad i = \overline{1,m}; \quad j = \overline{1,n}. \tag{4}$$

In this instance, the greater is the value S_{+j}, the more the environment of the object in question satisfies the positive criteria. The lower is the value S_{-j} (negative criteria of environmental factors), the more the environmental factors make negative impact on the object's utility degree. Anyway, the sum totals of the pluses S_{+j} and minuses S_{-j} of all alternatives are always, respectively, equal to all sum totals of weights of maximizing and minimizing criteria:

$$S_{+} = \sum_{j=1}^{n} S_{+j} = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{+ij}; \tag{5}$$

$$S_{-} = \sum_{j=1}^{n} S_{-j} = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{-ij}; \; i = \overline{1,m}; \quad j = \overline{1,n}. \tag{6}$$

This way, the calculations may be verified again. The relative weight (efficiency) of compared alternatives is determined considering relevant positive $S+j$ and negative $S-j$ features. The relative weight Q_j of each object aj is determined using the formula (7). Here $S_{-\min} = \min S_j$.

$$Q_j = S_{+j} + \frac{S_{-\min} \cdot \sum_{j=1}^{n} S_{-j}}{S_{-j} \cdot \sum_{j=1}^{n} \frac{S_{-\min}}{S_{-j}}}; \quad j = \overline{1,n}. \tag{7}$$

The process continues by identifying the priority of the objects in question. The higher is Q_j, the more efficient is the object—it has a higher priority. If $Q_1 > Q_2 > Q_3$, then the first object is the best. The above method is a rather simple way to evaluate and then to sort out the most efficient variants. The resulting generalised criterion Q_j depends, directly and proportionally, on the relative impact on the final result by the values x_{ij} and weights q_i of the criteria in question. Thus the result is an unbiased line of priority of the objects in question (Zavadskas et al., 1994).

Identifying the line of priority of the alternatives. The line of priority of the alternatives is determined considering their relative weights. The higher is the relative weight *Qj*, the more efficient the alternative is and thus gets higher priority. Potentially the best relative weight Q_{max} of an alternative will always take the highest position with other alternatives listed below.

Calculating the utility degree. The utility degree N_j of the object a_j marks the extent to which the object meets the requirements of the environment and stakeholders. The utility degree, therefore, helps to measure and justify the market value of the object in question. The more criteria show that the object meets the environmental conditions, the higher, proportionally, is the object's utility degree, which, in turn, has a positive impact on the market value (Kaklauskas 1999; Zavadskas & Kaklauskas, 2008). Then we proceed with identification of the weights, utility degrees and priorities of the environmental criteria that describe the objects in question. The efficiency degree E_{ji} of the object a_j is calculated. It shows the percentage by which the object a_j is either better or worse than the object a_i. E_{ji} is calculated by comparing the utility degrees of the objects in question:

$$E_{ji} = N_j - N_i . \tag{8}$$

The average deviation k_j of the utility degree N_j of the object a_j is calculated by comparing it with other objects (n-1).

$$k_j = \sum_{i=0}^{n} E_{ji} : (n-1). \tag{9}$$

The initial value of the object in question is calculated using the following formula:

$$x_{11} = \sum_{j=1}^{n} x_{1j} : (n-1). \tag{10}$$

The market prices of other comparable objects must be known if we want to calculate the market value of the technology or object in question in the above way. Generally, the analysis may be based either on the prices, say, determined by independent appraisers or analysts (when assets or investment projects must be analysed), or on the value estimated using economic methods (e.g., when technologies must be evaluated). The best way to frame the problem would be as follows: to determine, through an integrated analysis of the positive and negative features, which market value of the object a_1 would make it competitive in the market on equal footing with other comparable objects. The revised value $x_{11\text{-p}}$ of the object a_1 is calculated using the formula (11) :

$$x_{11\text{-p}} = x_{11} \times (1 + k_1 : 100). \tag{11}$$

It is determined whether the revised value $x_{11\text{-p}}$ of the object a_1 was calculated accurately enough:

$$|k_1| < S \tag{12}$$

here S is the accuracy in percent for the calculations of the market value $x_{11\text{-}R}$ of the object a_1. The market value $x_{11\text{-}R}$ of the object a_1 shall be deemed calculated when the following equation is satisfied:

$$x_{11\text{-}R} = x_{11\text{-}P} \tag{13}$$

The utility degrees and the revised market values calculated for the objects in question using this method depend, directly and proportionally, on an adequate criteria system, weights of the criteria and values of the weights.

If energy objects must be analysed or technologies selected, economic information and economic valuation approaches are insufficient to ensure objectivity. Multiple criteria analysis facilitates processing of extensive quantitative and qualitative information, analysing of a range of external value-affecting factors and assessing of energy production technologies with different dimensions in view of the most significant factors. Comparisons of the utility degrees based on the defining quantitative criteria alone would hardly be comprehensive and reliable in case of energy production technologies.

The utility degree and the market value of the object, project or technology in question may be measured using the following sequence (Fig. 2):

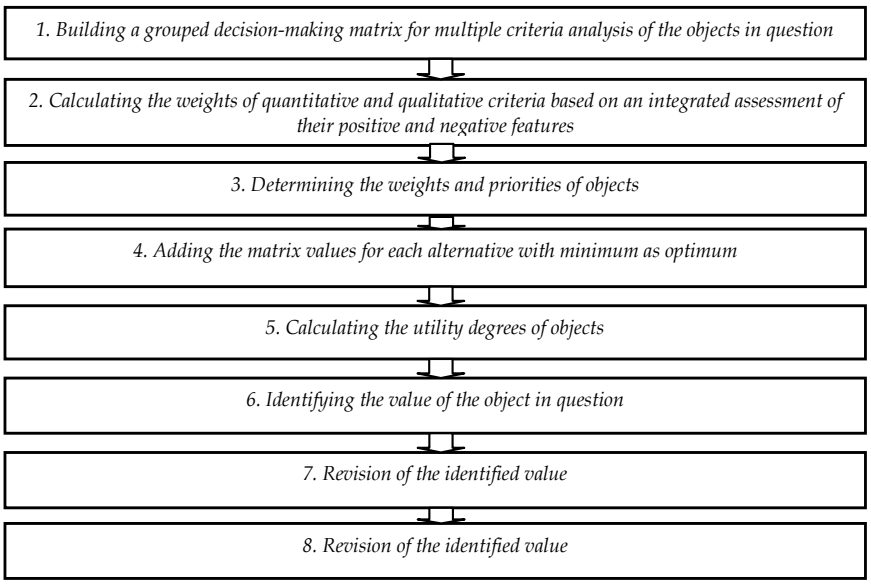


Fig. 2. The algorithm for calculation of the market value

2.1 Methods for identification of criteria values and weights

It is expedient to make an integrated analysis of objects by combining quantitative and qualitative assessment. Multiple criteria analysis methods are the right tool: they allow a more thorough look at an object through combined quantitative and qualitative criteria; they help to make an analysis by integrating both technical data and the results obtained using economic valuation approaches or analyses. The so-called *weighting methodologies* that help to determine criteria weights are one of the methods to determine preferences, and to make quantitative and qualitative assessment of the weights attributed by people to some criteria, which are then used to make decisions on implementation of a certain instrument. These methods encompass a wide range of surveys, which help to approximate the quantitative expression of social values, and a wide range of indirect valuation approaches, in which significant factors are identified through debates in task forces or between stakeholders by

identifying their preferences. The Kendall's coefficient of correlation (Kendall, 1970) is the most popular tool in assessments of survey results.

The methods used to identify the values and weights of criteria that define environmental factors may be classified conditionally into two groups:

- methods for expression of criteria in quantitative terms (e.g. expression in money, technical parameters).
- methods for expression/description of criteria in conditional measures (scores, percentages). Such criteria are qualitative.

While quantitative criteria may be expressed in money or technical parameters for technologies, the measurement of the values and weights of qualitative criteria is a more complex task. Weights for qualitative criteria may be determined through analysis, scientific studies and databases, by comparing equivalents, and by analysing macro, meso and microenvironment in regions with similar development degrees or development trends. Values and weights of qualitative criteria are most often identified through expert or sociological surveys.

The expert assessment must necessarily include identification of criteria weights; otherwise it is difficult to assess the reliability of the research. Such calculations aim to determine whether different experts agree to a sufficient degree for expert assessment results to be a reliable basis. Kendall's coefficient of concordance W is the measure that describes the degree of agreement between expert opinions. This coefficient is calculated using a specific table of expert assessment indicators and the formulae available in literature sources. The calculated coefficient W shows the degree of agreement between individual opinions. This coefficient will be equal to 1 if all experts share one opinion. When all ranks differ and expert opinions do not match, the coefficient will be equal to 0 (Kendall, 1970; Zavadskas et al., 1994).

It is often difficult to reconcile opinions on most economic decisions, particularly in the energy sector. Reliable research demands for choosing experts from such social and interest groups, and with such knowledge and qualifications in the respective area, as to make it possible to reconcile the opinions on qualitative indicators. Decisions on the development of the energy sector, on choice of technologies, their performance and application, encompass a multidimensional and conflicting task: to minimise costs, to minimise the effect on environment, to ensure reliable energy supply, to supply more energy, to ensure socially-responsible pricing, to develop renewable resources, etc. Strategic decisions on the development of the energy sector, therefore, must be based on multiple criteria analysis or multiple criteria evaluation, because it helps to consider the relative importance of criteria to the decision-maker thus reconciling political, economic, environmental, social and other criteria, and to select the best solution in view of all criteria. *Multiple criteria* methods for evaluation or decision-making are, therefore, an attempt to simultaneously assess several alternative solutions on the basis of a set of contradictory criteria.

Multiple criteria analysis methods are, first and foremost, applied to determine the preferences of stakeholders involved in the decision-making process. These methods are of particular importance now, when public participation is part of all evaluation procedures dealing with strategic assessment of the impact on the environment. Community consulting is the best way to ensure that national policies serve the public. The energy sector is of particular relevance, because it is strategically important to each citizen both in terms of economic and social welfare. It explains increasing popularity of multiple criteria evaluation methods in decision-making on important issues of energy development.

3. Building the model for analysis of environmental factors in a multiple criteria decision support system

An important aspect in any analysis of alternative decisions is selection of suitable evaluation criteria, and indicators for quantitative and qualitative assessment. It is also important to build systems of criteria that define alternatives best. A multiple criteria task, therefore, involves research of environmental factors, as well as identification of the most important factors that affect activities or decisions. Lately, multiple criteria evaluation methods are also widely used in sustainability assessments. In the energy sector, decisions are made at three levels: decisions related to development scenarios in the energy sector or strategic priorities of national energy development (macro level); decisions that reconcile the sector's role in the national economy, the processes within the sector, the effect of the processes on the environment and the legal environment that regulates the sector (meso level); and decisions on the choice of specific technologies for energy production or implementation of specific energy projects (micro level). The analysis of alternative decisions made at these different levels demands for relevant criteria for the assessment of such alternatives, as well as for indicators to be used in quantitative and qualitative evaluation of the alternatives. Problems involving multiple criteria highly depend on the compiled decision-making model.

3.1 Analysis of environmental factors and stakeholders

Expert methods are used in environmental research to determine the factors of various levels, as well as systems and subsystems of their defining indicators, which provide a thorough description of the activities related to the sector in question. The following components of the developed model for analysis of environmental factors make the biggest impact on its effectiveness and performance:

- macro, meso and microenvironment with the defining factors;
- groups taking part in the decision-making process and their chances to influence the decision.

Macroenvironment Factors. This factors define the level of national or industrial performance. Besides, macrolevel factors affect the development degree of separate industries. The performance of energy companies vastly depends on the integrated effect of macrolevel variable factors, such as national economic, political and cultural development, international and political commitments, agreements, legal acts and strategies, the market, the tax system, conditions in the loan market, inflation, dependence on natural resources and raw materials, etc. The changes in the performance of the industry in question—decreasing or increasing demand for energy resources—depend on the integrated effect of macrolevel factors.

PESTEL (*Political Forces*, *Economical Forces*, *Socio-Cultural Forces*, *Technological Forces, Environmental Forces, Legal Forces*) is the most suitable tool for the research of macroenvironment in energy sector. This analysis covers the main aspects of macroenvironment, namely political, economic, socio-cultural, and technological aspects, and, most importantly, it also includes the environmental and legal aspects. This analytic technique is widely used in practical applications, for instance, analysing the necessity to change the efficiency of energy consumption in Chinese regions, at the same time considering the impact of environmental factors—the links between a variety of political, economic, social, technological, environmental factors aspects and legal acts, as well as the main factors (Shilei & Yong, 2009). The analysis of macroenvironment must include a

thorough analysis of political and legal environment, because the energy sector is subject to strict legal regulation. The activities are regulated by EU and national legal acts. Recently, EU member states started harmonisation of these legal acts and their transfer into national legislative bases. This process simplifies the analysis of legal environment. The aspect of the environment protection is equally important. The PESTEL analysis of the environment includes quantitative (extrapolation, mathematical modelling, etc.) and qualitative (scenarios, Delphi, etc.) forecasting methods. The main factors for PESTEL analysis in the energy sector are shown in Table 1.

Factor	**Components**
Political	EU enlargement, tax (excise) policies, agreements on development strategies, etc.
Economic	national income, inflation, unemployment, industrial development, etc.
Social	demographic features, emigration, income distribution and levels of consumption, supply of skilled professionals, the culture of industrial relations, etc.
Technological	innovations, development of new products, speed of technology updates, dissemination of technical know-how, etc.
Environmental	global warming, pollution reduction and environment protection
Legal	laws that promote competitiveness in the sector, healthcare and labour laws, etc.

Table 1. The main factors for PESTEL analysis in the energy sector

Mesoenvironment Factors. The analysis of mesolevel environment is oriented towards the goals of a specific sector, its role in the national economy and the industry, the features which shape the type of activities, profit, processes within a specific industry, the impact of the processes on environment, fulfilment of the sector's social role, documents regulating specific activities and relations with national authorities. It is an intermediate level between microeconomics and macroeconomics (Pillet et al., 2005). The analysis of mesoenvironment aims to look into the relation between the environment of the object in question and the economics. This environment is analysed considering such factors as institutions involved in legislation of a variety of legal and normative acts at various levels, in supervision and in control. A direct relationship links the decisions made by such institutions and the legal acts which regulate corporate activities with corporate plans, and decisions. Indicators that look into the ability of a specific object to achieve its economic goals in a specific legal environment and to handle the environmental issues through resource-saving manufacturing tools and increasing use of renewable energy sources are significant in the analysis of mesolevel factors. It is at the mesolevel that the environmental dimension, and the external effect (effect of by-products and pollution on the environment) of activities, is analysed in the energy sector or in separate energy production technologies. An important aspect of this level is public participation in issues related to the quality of living and work environment. Another important aspect, though rarely considered, is social responsibility of energy companies. This aspect must be considered not only in the analysis of specific corporate activities, but also when drafting strategies, selecting development scenarios and technologies, because the activities in the energy sector are important not only from the economic perspective, but also when dealing with a range issues relevant to the public—

related to environment protection, resource saving, readiness to introduce innovations, etc. It expands the boundaries of macro and mesoenvironment, as well as the effect of these factors on the performance in the energy sector.

Environmental Factors. Making an analysis of activities in the energy sector, it is worthwhile to make a more thorough assessment of environmental factors. Companies operating in this sector make a considerable impact on the environment. Organic fuels of limited quantities are widely used in the process of energy production. Environment is polluted by SO_2, CO_2, NO_x and other types of particulate matter, which are a by-product of energy production and can affect the soil, the water, the air, the biological cycle, and can generate huge amounts of hard waste (Ashina & Nakata, 2008; Streimikiene & Pusinaite, 2008). Despite high economic performance parameters, nuclear energy includes a complex and expensive burying of radioactive waste accumulated during the energy production cycle. Even electricity transfer through open high voltage lines generates electromagnetic fields, the effect of which is considered in legal acts regulating operation of such objects. Cleaner production tools, therefore, is a very effective and economically beneficial course of action, when we must analyse the sector's development and select the energy production technologies. For many years, EU countries apply environmental protection measures based on market factors, for instance, environmental taxes to increase the market share of products, processes and services which are more acceptable in terms of environmental protection. Such taxes encourage companies to allocate more funds for R&D and to invest into technologies less damaging to the environment or consuming fewer resources (Staniskis & Stasiskiene, 2006).

Microlevel Factors. Microenvironment factors are related to a specific company that uses the relevant energy production technology, they affect its ability to achieve its goals. These factors embrace all things related to value delivery to the customer: activities of the company, of its suppliers (from primary energy sources to supply, distribution and service companies) of its competitors, consumers and the public. These factors depend on macro and mesolevel factors. The energy sector must continuously keep high levels of infrastructure maintenance, must modernise facilities, must expand, and must introduce innovative technologies and management processes. The efficiency of the sector's development and implementation of investment projects is affected by various microlevel factors: land prices; lengthy procedures of territorial planning; efficiency of the process related to reconstruction, modernisation, and supply of technologies, mechanisms and equipment; financing conditions of development projects; etc.

Analysis of Groups that Influence Decisions. The analysis of environmental factors cannot be thorough before the stakeholder groups, which affect activities and decisions, are considered in assessment of specific environment in the energy sector. Activities in this sector are controlled and coordinated by national institutions and a variety of EU institutions. Various institutional participants—starting with international alliances, committees of associations and ending with trade unions—have a direct influence on sector's activities. Assessment of stakeholder groups considers the type of their influence, their expectations, requirements, represented institutions, business sector, and possible effect on the decision. All macrolevel stakeholders are active in the energy sector: national governments, local governments, the public, suppliers of resources and technologies, manufacturers, nature activists, etc. In the energy sector, the same stakeholder group may have contradictory interests. For example, residents usually support companies which use renewable resources but are against construction of wind parks in their neighbourhood (Sims et al., 2008), suppliers of raw materials are interested in the development of the

thermal energy sector and challenge the development of nuclear energy. The analysis of stakeholders and the obtained results help to assess the requirements and expectations of various groups, to evaluate them and to search for ways to influence hostile groups or to assist and strengthen the supporters. The interrelations of stakeholders are shown in Fig. 3.

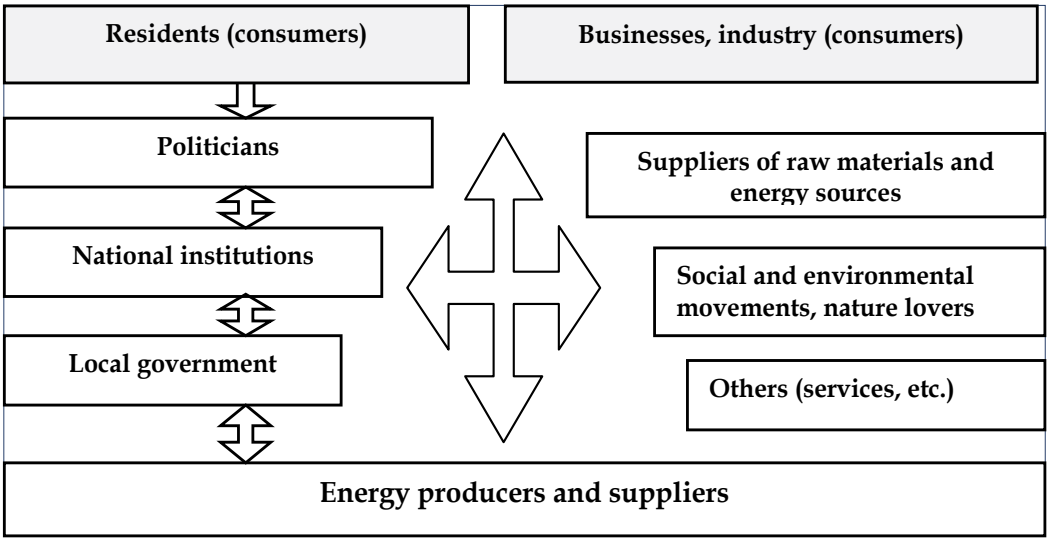


Fig. 3. Interrelations between stakeholders and the organisation in question

Consumers are also important members of the energy sector. Growing prices of raw materials, as well as electricity production, distribution and supply, make various groups observe the processes in the energy sector and participate in the development of strategies and in decision-making. Active involvement of stakeholder groups and political organisations may lead to economically unreasonable decisions harmful to the environment protection, but it also can and does make a positive impact: promotes transparency and responsibility, affects the process of market liberalisation.

3.2 Model for analysis of environmental factors in the energy sector

The above-discussed analysis of environmental factors helps to formulate the model for assessment of the energy sector (technologies), which is shown in Fig. 4. The model integrates the key factors affecting the performance and the requirements set forth in environmental policies. The model may be implemented through selection of decision support instruments, through comparison, ranking and assessment of energy production technologies.

Environmental analysis in the energy sector starts from identification of the environment factors with the biggest effect on performance; then a system of quantitative and qualitative value-affecting criteria must be built and their weights identified. The environment of separate energy objects or technologies must be described considering the environment of the entire sector and relevant information that affects activities must be selected. A most expedient variant is to make use of indicator systems built for the analysis of sustainable development and of criteria sets suggested by various authors, if the sets suit the specifics and environment of the sector in question. The most efficient environment of the objects in

question is determined by comparing the values and weights of environmental criteria and by analysing conceptual information. The impact of the environment in question may be described only having a system of numerous criteria with different meanings and dimensions.

Having selected the methods of multiple criteria analysis and collected quantitative and qualitative information, it is now time to identify the values and weights of criteria. Multiple criteria analysis of alternatives helps to select the most efficient energy production technology in terms of environmental factors. The developed model for evaluation of environmental factors is handy in selection of decision support instruments with emphasis on the policy of sustainable energy development. If an energy development plan is required, the assessment, comparison and ranking of energy production technologies in the context of sustainable development helps to select the most promising projects, and the best future technologies of energy production, also to include them in energy models. Assessment of scenarios related to environmental policy instruments helps to select the environmental policy instruments that secure achievement of the priority goals related to the energy sector development at the lowest cost for the public.

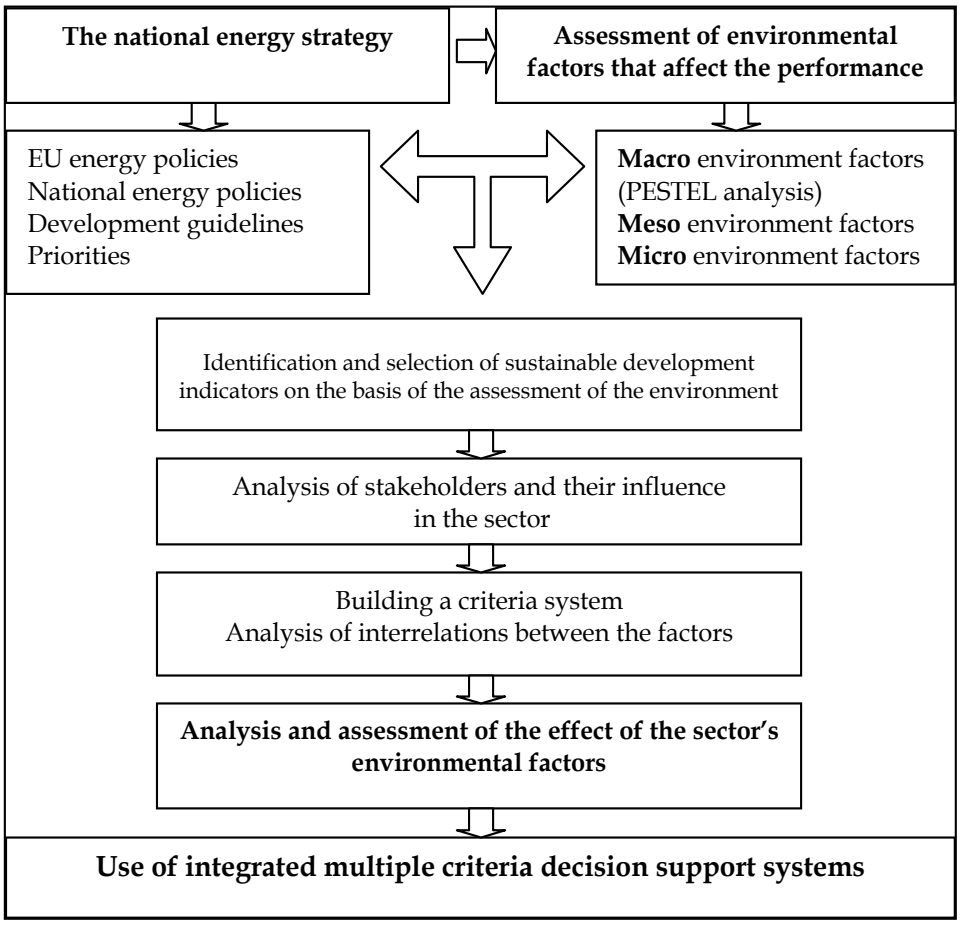


Fig. 4. Model for analysis of environmental factors in the energy sector

4. Decision making in the energy sector with the help of multiple criteria evaluation methods

A decision support system is treated as a chance, in view of the priorities, to select the best alternative from a set of alternatives framed or offered by the system. A typical decision-making procedure includes four main phases (Booty & Wong, 2010; Bergey et al., 2003):

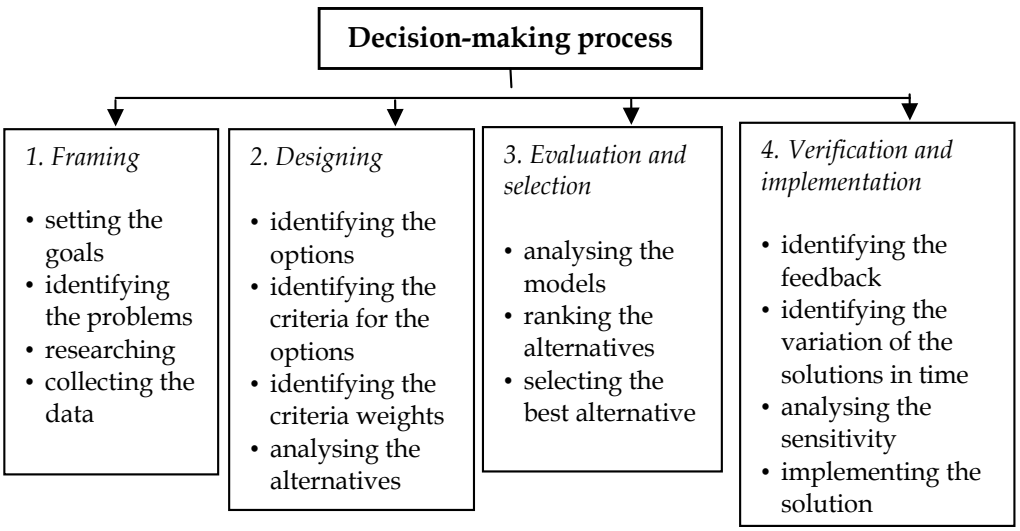


Fig. 5. Phases in decision-making

Each decision has a different context, thus we must consider:

- the goals;
- the object in question and its environment;
- the criteria defining the efficiency of our choice;
- the variables of the solution;
- the restrictions and risks.

The decision-making process starts from *goal setting* and problem identification. When our decision-making concerns the energy sector, the first and foremost step is defining the development priorities or the strategic goals of the development in the energy sector. The goal setting is of vital importance to make right decisions. The strategic goals must be clear and have explicit wording, they must be specific, measurable, matched, realistic, and time-dependent. Sometimes it is best to classify goals by their level. For instance, the Green Paper separates primary, interim and direct goals (Commission of the European..., 2006), but in real life primary and direct goals are more frequent. Primary goals are usually defined by variables at the strategic or higher level (degree of economic growth, social cohesion or sustainable development). These goals may be set forth in the White Paper, national strategies of economic development, and in other national strategic documents of importance. Direct goals are directly related to political instruments, programmes, or projects. Considering the suggested goals one must focus and select the criteria which contribute to direct—urgent—goals. These goals are seen in national energy strategies or development programmes for national energy sectors.

The next phase—*designing*—includes setting of the options, which may be handy while seeking the set goals. The options may include ranking of energy development scenarios or

selection of the principal key political instruments, such as new environmental taxes for the energy sector, or individual projects to improve the reliability of energy supply. The most promising options are to be developed further. They may include the key political plans, such as changes in the tax policy, or more detailed options related to preparation of individual investment projects and development scenarios for electricity with the lowest costs. Here, feedback is important. The designing stage also must choose the way to compare the impact of different options or alternatives on implementation of the goals of the first priority. The selected criteria must reflect the performance related to the goals. Each criterion must be measurable for the sake of assessment—it must be possible to pick specific indicators that will help to assess the impact of choices on achievement of the set goals. This assessment may be also qualitative; it may express the relation between a certain criterion and the selected option. Identification of weights is an important aspect; a range of methods (described above) may be used for that purpose and the process also includes the assessment of the preferences of social groups.

The *evaluation and selection phase* includes the analysis of the alternatives. Practical solutions most often require financial analysis, price efficiency analysis, and, in some fields, the cost-benefit analysis. These types of analysis fully, or partially, depend on the expression in money. Conversely, multiple criteria analysis is directly related to some uncertainties and includes qualitative assessment; it allows to make the aspects considered sensitive by the public part of the decision-making process. The most solemn part of this phase is ranking of the alternatives to identify the best solutions. The reliability of the end result depends directly on the identified issues, analysed data, selected system of indicators, thoroughness in the assessment of environmental factors, etc.

The *implementation and verification phase* is for implementation of the selected solution and, if possible, verification of the result. It includes identification of the feedback and the variability of solutions in time, as well as the sensitivity analysis.

4.1 Information technology in decision-making: automated decision support systems

Multiple criteria analysis methods help to assess alternatives by processing huge amounts of information and data. Information technology may facilitate the process and make it suitable for problem solving at various levels. Mathematical data processing methods are used to build an information processing system for automated handling of any problems related to multiple criteria analysis of variants using a developed algorithm for the analysis of environmental factors. Automated decision support systems have a fundamental advantage: they may be modelled and adjusted on the basis of other systems, considering user needs and specific features of the alternatives in question. Automated decision support systems have the following advantages:

- possibility to base decision-making on data provided in a form adequate for the problem in question: numbers, text, graphical expressions, formulae, etc.;

decision support systems may be modelled and used both for individual and for group decision-making;

- customised software enables processing of huge amounts of information, and ensures access to data from diverse sources, as well as selection and addition of data that best reflect the needs and requirements;
- integrated software facilitates development of data models and complex studies, preparation of reports for users via a range of channels (internet, e-mail, printing and mobile devices).

Quite a few decision-making methods, expert decision support systems, neural networks, spreadsheets and analysers to facilitate decision-making are currently available worldwide. Decision support systems, by the type of support, are classified as follows: 1) individual decision support systems (IDSS); 2) group decision support systems (GDSS); 3) negotiation support systems (NSS), and 4) expert systems (ES). Expert systems are intended as a tool for skilled professionals of certain fields, for experts. Therefore, expert systems must incorporate a range of indicators that summarise the specific field-related knowledge and comprehensive information that describes the issue in question. The system "recognises" the situation, identifies the diagnosis, gives questions, and recommends decisions. These systems have many secondary functions: they make questions, model alternative decisions, and offer conclusions and recommendations.

Expert systems are gradually becoming an inseparable part of any decision-making process. Expert decision support systems are widely used in environmental studies, in political decision-making, in evaluations of strategies, development plans and projects, and in selection of alternatives, when such selection is not easy due to abundant factors that may affect the decision.

Expert systems may be classified into two groups:

1. Working expert systems: they are elaborated and approved expert systems continually used in decision-making and undergoing continual improvement (Svensson, 2002; Masood & Soo, 2002).
2. Conceptual expert systems: any systems in development and any expert systems for initial scientific assessment (Benavides & Prado, 2002).

Decision support systems successfully helped to handle a multiple criteria problem in the energy sector. One decision support system was developed specifically for Ghana and used in environmental research as a tool to justify the decisions related to restructuring of the electrical power industry (Bergey et al., 2003). One application was developed as a tool to analyse the dynamic price changes in the US wholesale market of electric power. This software is used for practical applications and is an effective instrument in modelling and simulation of situations in the market of electrical power trade (Sueyoshi & Tadiparthi, 2008). Literature sources cite a multiple criteria problem that was handled selecting competitive nuclear technologies, assessing environmental factors, and performing the analysis of alternatives. This decision-making process was based on the AHP methodology, which was also the basis for the relevant decision support system (Deok & Johoo, 2010; Shen et al., 2010). Decision support systems are used to select and maintain renewable energy technologies (Yue & Grant, 2007; Sliogeriene et al., 2009). They are also used to analyse the energy policies of EU member states, to set uniform evaluation criteria that help to consider countries with different levels of development and to assess their energy development in line with the principles of sustainable development, to assess the expansion of renewable energy sources, environment protection, and the performance of the energy sector (Patlitzianas & Psarras, 2007; Drozd, 2003). The ELECTRE methodology was used to develop a decision support system, which analyses the instruments that can help to improve the efficiency of energy consumption in view of environmental factors. Decision models also incorporate the cost-benefit analysis thus enabling the decision-maker to verify the final outcome of each decision (Neves et al., 2008). More decision support systems based on a number of multiple criteria analysis methodologies are available. But all these systems share one advantage: they can process huge amounts of information, and assess how the dynamically fluctuating impact of the environment affects the decisions.

Today the economic community of countries is becoming integrated fast, morphing into a single economic system. National economies become cells of a global economy, thus diverse decisions related to activities and development must be considered within a broader context. Integration into the community of global economic structures demands for internationally recognised and universally intelligible analysis and assessment methods, as well as for universally intelligible requirements and assessment criteria. Use of decision support systems can help to handle such tasks, because the systems facilitate use and assessment of a wide range of information and data, to come up with criteria systems for relevant factors, and to suggest solutions in line with the goals of the international community. Globalisation demands for knowledge how decision-makers come up and make decision in other parts of the world and how information systems (IS) may facilitate decision-making. Analyses consider the differences between national cultures, values, decision processes, and decision-making. "The existence of international differences in analyzing and conceptualizing strategic decisions raises doubts about the global applicability of IS such as decision support systems and executive information systems. The success of knowledge management and information systems in different countries and cultures will depend critically on how well IT applications are adapted to the decision styles of their users" (Martinson & Davison, 2007).

Many tasks related to economic development are multiple criteria tasks. Within the context of sustainable development, energy is in the spotlight. The costs related to the building of energy infrastructure, exports and imports of energy sources and refined products, and end-user energy prices have enormous impact on the development of national economies; therefore, it is the trends of energy-related economic indicators that point out the potential of country's economic development. Typical economic valuation methods and decision validation methods used in the market are the usual choice when one comes to practical issues of the development of the energy sector and selection of energy production technologies. These methods lack tools for integrated assessment of environmental factors, which could determine selection of more efficient and more economically promising technologies, as well as better environmental solutions.

The energy sector is marked by its versatile aspects, uncertainty, and the influence of interests; therefore, to validate various operational decisions, to model the development scenarios, and to make the decisions objective, we need methods, which would facilitate use of several interacting indications of values, validate the decisions in the decision-making process, and make the process more transparent. When it comes to modelling of variants related to the development and use of energy production technologies, as well as to management decisions, decision support system, which combines changing factors of macro-, meso- and microenvironment, environmental factors, as well as economic and technological indicators, would facilitate handling of practical tasks related to management of energy systems, to selection of effective technologies, to analysis of prices, and to search for the best development or management solutions.

4.2 Preparation of data for an automated decision support system

The selected objects operating in the energy sector were analysed using the method for multiple criteria complex proportional assessment and multiple criteria measuring of the utility degree and the market value, namely the COPRAS method. This method was selected because it allows:

- identifying the utility degree of alternatives, which shows the percentage at which one alternative is either better or worse than other alternatives in question;
- identifying the priority of alternatives;
- identifying the market value of alternatives in question;
- integrating classic valuation approaches based on economic indicators in the analysis.

The environmental studies performed by the authors were used as a basis to build the Decision Support System for Measurement of the Effect by the Environmental Factors on the Value of Energy Companies; the system, after more thorough studies and corrections, may be suggested for use in the energy sector as a tool helping to validate diverse decisions, to analyse and select energy production technologies.

In multiple criteria analysis methods, energy production technologies, or the companies which use such technologies, must be assessed considering quantitative (operational territory, number and length of engineering infrastructure objects, technical and technological parameters, economic indicators) and qualitative (condition, degree of modernisation and new technologies, environment protection, political, legal and legislative restrictions) criteria of the current market conjuncture, which describe the object, as well as other indicators that affect the value. When the key factors that may affect the result of the problem in question are selected, it is time to build the set of criteria and to determine weights of environmental indicators pertaining to the object in question.

When we have to assess alternatives in the energy sector, and to determine their efficiency, analysis focuses on a variety of energy production technologies, which differ both by their qualitative and quantitative parameters; the method of integrated analysis, however, allows to separate the factors that affect value, to determine the evaluation criteria, and thus to compare such technologies, to calculate their utility degrees, to make a priority line and to calculate the revised market value. The method of integrated analysis includes the following main steps:

1. Identification and description of qualitative and quantitative criteria that affect activities of a technological complex for energy production;
2. Building of an integrated database based on the description of the objects in question;
3. Use of multiple criteria analysis methods as a means to determine the utility degree and to revise the market value of the alternatives.

When the objects are already described in both quantitative and conceptual forms, an integrated database must be built to describe in detail internal and external factors that affect the value of the objects in question. The database is a basis for multi-variant designing and multiple criteria analysis of the objects. Even though the amounts of available data and information are huge, handling of a multiple criteria analysis problem is made considerably easier by using automated intelligent decision support systems designed in view of the selected goals.

4.2.1 Building of criteria systems

In order to assess the macro, meso and microenvironment factors that affect the utility degree, and the revised market value of the technological facilities for energy production, as well as in order to compare the objects, a system of defining criteria must be built. The system is built so that it helps to analyse the environment of the selected technological facilities together with economic and technical indicators pertaining to the objects. We divide the system of criteria into two main groups: qualitative criteria and quantitative criteria. The groups are then subdivided into subsystems, which may, in turn, be subdivided further:

Qualitative criteria:
- the subsystem describing the impact of macrolevel factors;
- the subsystem describing the impact of mesolevel factors;
- the subsystem describing the impact of microlevel factors.

Quantitative criteria:
- the subsystem describing the technical data;
- the subsystem describing the economic data;
- the subsystem describing the preliminary value.

The integrated method for the measurement of weights is used when the objects may be assessed using a sufficient pool of relevant quantitative criteria. A system of quantitative criteria includes economic, technical, or other criteria that describe the objects. Energy production technologies may be described by the rated capacity of the objects, the volume of energy production, the cost of energy production, profitability, number of users, book value or replacement value of the technological facilities and assets, etc.

Qualitative criteria that describe the impact of the environment are by no means less important than the quantitative criteria. The choice of qualitative criteria depends on the goals of the task; such criteria may describe the economic context and priorities, as well as attitudes and expectations of a range of social groups. Expert methods are used to identify the weights of qualitative criteria. When the integrated method is used to identify the weights of qualitative criteria, quantitative and qualitative features are considered.

4.2.2 Identifying values and weights of criteria

1. Expert assessment of energy production technologies (companies). The objects of the Lithuanian energy sector selected for the multiple criteria analysis problem to be handled with the COPRAS method produce energy using different technologies, and different primary energy sources. The objects are:

- *Kruonis Pumped-storage Hydroelectric Power Plant* - uses a combination of technologies based on renewable sources and traditional technologies;
- *Kaunas Hydroelectric Power Plant* - uses only technologies based on renewable sources;
- *The Lithuanian Power Plant* - uses traditional technologies (fuel oil and natural gas as the primary energy sources). Closing of Ignalina Nuclear Power Plant made this power plant the main electricity producer in Lithuania.
- *Experimental geothermal power plant* - uses a combination of technologies based on renewable sources and traditional technologies.

A questionnaire was compiled as a tool to analyse and determine the weights for qualitative criteria of the selected technological facilities of energy production. The questionnaire was compiled making it suitable for assessment of the main qualitative criteria defining the selected technological facilities of energy production. The questionnaires were filled in by groups of experts representing diverse social groups with different interests. The questionnaire was divided into three subsystems, each dealing with the effect either of macro, meso or micro qualitative factors. The subsystems include the criteria that describe the environment of the objects in question. Conditional measures—scores between 1 and 10—determine the criteria weights. Experts attributed bigger weights to the criteria they considered more important and such criteria had bigger impact on the final results of the assessment. The average estimates for each criterion determined on the basis of the results of expert assessment, and the resulting weights of the criteria, were identified for each group of experts.

Identifying Values and Weights of the Criteria. The results of the expert assessment were used to determine the weights of criteria defining the objects in question, and to arrange the weighted criteria in the order of their priority. Selected groups of experts took the job to determine the values and to list the criteria in line with the priorities selected by the experts. When the weights and the priorities of the criteria were clear, it was time to determine the key environmental factors that affect technological facilities of energy production. The research included a total of 29 criteria assessed by six groups of experts. The reliability was ensured by assessing the agreement between expert opinions with the help of the Kendall's coefficient of concordance. The value of the coefficient of concordance *W* is 0.29, which is above zero and thus ensures sufficient reliability of criteria weights obtained by the outranking method. The result, however, shows low degree of agreement between expert opinions. It is only natural when the experts selected for such expert assessment have different attitudes towards the environment that affects this industry. Moreover, 29 criteria selected for our research make it more complex and the deviation of expert opinions is more likely. To make the research more reliable, it is possible to determine criteria weights for each subsystem of the analysed environment—macro, meso and microenvironment—separately. This would make the research simpler for experts. Expert assessments are stochastic: changes in the composition of the groups would also change the assessments of the indicators, which determine the coefficient of concordance (Podvezko, 2005).The ranking of the key criteria obtained after the expert assessment is shown in Table 2.

x_i	**Criterion**	**Weight**
24	Experience of CEOs	0.061
14	Profitability	0.055
25	Supply of skilled professionals	0.053
27	Readiness to choose and introduce innovations	0.052
1	EU regulation of activities related to electric power supply	0.049
10	Environmental regulation	0.049
6	Investment conditions	0.047

Table 2. Criteria line-up by weight after the expert assessment

The results of expert examination suggest that experts consider the experience of CEOs, profitability, supply of skilled professionals, and readiness to introduce innovations as important factors for different technological facilities of energy production. The experts believe that environmental regulation, technological shift, and corporate social responsibility are also of importance. The latter criterion was ranked 12th of 29. It shows that both employees who were members of the groups of experts and other experts from energy companies expect responsibility when it comes to activities, to handling of environmental issues, and to the response to public needs. Such criteria as competitive environment and relations with authorities received low weights. This is because technological facilities of energy production are currently monopolies in the energy sector, the companies that could compete do not yet have favourable conditions to enter the market.

4.3 Decision support system for measurement of the effect by the environmental factors on the value of energy companies—ESIAPVN-DS

This section presents a case of the integrated multiple criteria analysis of alternatives using the Decision Support System for Measurement of the Effect by the Environmental Factors on the Value of Energy Companies. The system was developed using the algorithm developed by the authors and following the methodology suggested by the scientists E.K. Zavadskas and A. Kaklauskas (Zavadskas et al., 1994; Kaklauskas, 1999; Zavadskas & Kaklauskas, 2008). The system may be used to measure the utility degree, the priority and, if required, the market value of energy production technologies with different technological and economic parameters. The system has an innovative feature: recommendations and other information to facilitate decision-making are provided after the analysis of each criterion describing the environment. Environmental research in the energy sector and data assessment with multiple criteria analysis methods helped to build a universal set of criteria defining the environment in question.

The experimental Decision Support System for Measurement of the Effect by the Environmental Factors on the Value of Energy Companies (ESIAPVN-DS) in designed for measurement of the utility, priority and value in energy companies which use different energy production technologies. The system is suitable both for analysis of separate energy production technologies and of technological facilities (companies). The ESIAPVN-DS system speeds up handling of tasks, provides rather accurate and unbiased results, and enables the access to comprehensive data about the object in question. A huge advantage in the system is the interim result, which shows the impact of each criterion on the utility degree and value. In this research, this decision support system is used for the first time to perform multiple criteria analysis of energy objects with different operating parameters, features and levels of external effect. The decision support system consists of the following models:

- measuring model for the initial criteria weights;
- model for multiple criteria analysis and priority setting of objects;
- measuring model for the utility degree of objects;
- measuring model for the market value;
- recommender model.

Users, with the help of the model base management system, may choose any model they need. The system is designed in a way that the output of certain models is used as the input in other models, while the output of these latter models is used as the input yet in other models. The ESIAPVN-DS system may process huge amounts of data to solve the main task. The system is, however, user-friendly.

When the ESIAPVN-DS system is used for measuring of the market value and the utility degree, the procedure includes the following stages:

Stage 1. Preparation of the data about the objects in question (projects, scenarios, technologies, technological facilities). It involves analysis of the materials describing the specific environment of the objects in question, the factors that affect the activities. The objects in question are described in quantitative and conceptual forms;

Stage 2. The system of quantitative and qualitative criteria built on the basis of the analysis of the factors that affect the environment of the object in question is used to determine the criteria values. A group of experts determines the initial values of the criteria, while the weights are already known from the expert assessment method. When the ESIAPVN-DS system and the built set of criteria are used, there is no need to determine the weights anew;

Stage 3. Input of the expert assessment data into the ESIAPVN-DS system for data processing;
Stage 4. The system's model bases are used to process the data: the programme compiles a matrix of normalised weighted criteria expressed in numbers, and determines the revised market value, priority and utility of the objects;
Stage 5. The ESIAPVN-DS system's recommender model analyses the results: the system gives the values of the criteria that affect the value, and automatically prepares and suggests recommendations on ways to reduce or increase the impact of certain criteria and thus to increase the market value of the object.
The ESIAPVN-DS system is available online at the address <*http://iti.vgtu.lt/elektra/*> . In the main page of the ESIAPVN-DS system, the system's administrator or a user may log in and fill in the main data tables (Fig. 6). The figure shows the expert assessment results of the energy objects selected for our analysis (Kruonis Pumped-storage Hydroelectric Power Plant, Kaunas Hydroelectric Power Plant, the Lithuanian Power Plant and the experimental geothermal power plant).

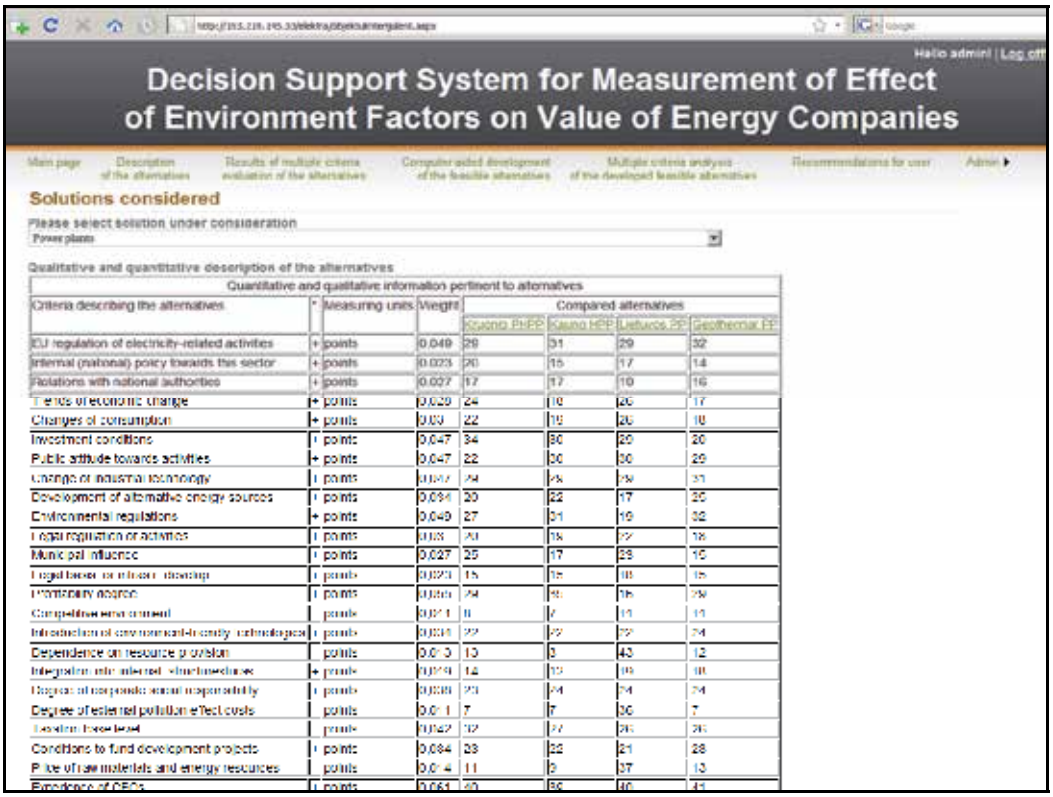


Fig. 6. The main window of the ESIAPVN-DS system with values of the environmental criteria that affect energy companies

The system automatically processes the data and, upon clicking on relevant links, shows the results of multiple criteria evaluation of the objects. The system automatically solves the multiple criteria task: performs multiple criteria analysis of the selected objects, measures their utility degree and market value and thoroughly analyses the impact of the criteria on

the value. Simple active links in the main window lead to the model required for your decision. To access the measuring model for the initial criteria weights you must click "Description of the alternatives". This link leads to the expert assessment of the criteria and to their weights (qualitative and quantitative criteria, their values and weights are shown for each object). A click on the menu item "Multiple criteria analysis of the developed feasible alternatives" in the ESIAPVN-DS system leads to building of a normalised decision-making matrix, which helps to assess the criteria and to determine their values and weights.

The next link in the main window, "Results of multiple criteria evaluation of the alternatives", leads to the automatic assessments of the objects in question based on the criteria system (Fig. 7). The system shows the numerical values of normalised weighted

Results of multiple criteria evaluation of the alternatives

Please select solution under consideration

Power plants

Quantitative and qualitative information pertinent to alternatives							
Criteria describing the alternatives	*	Measuring units	Weight	Compared alternatives			
				Kaunas PHPP	Kaunas HPP	Lietuvos PP	Geothermal PP
EU regulation of electricity related activities	+	points	0,049	0,0117 AVG MIN	0,0126 AVG MIN	0,0117 AVG MIN	0,013 AVG MIN
Internal (national) policy towards this sector	+	points	0,023	0,0057 AVG MIN	0,0052 AVG MIN	0,0059 AVG MIN	0,0049 AVG MIN
Relations with national authorities	+	points	0,027	0,0067 AVG MIN	0,0067 AVG MIN	0,0074 AVG MIN	0,0060 AVG MIN
Trends of economic change	+	points	0,028	0,0079 AVG MIN	0,0069 AVG MIN	0,0066 AVG MIN	0,0050 AVG MIN
Changes of consumption	+	points	0,03	0,0078 AVG MIN	0,0067 AVG MIN	0,0092 AVG MIN	0,0054 AVG MIN
Investment conditions	+	points	0,047	0,0141 AVG MIN	0,0125 AVG MIN	0,0121 AVG MIN	0,0080 AVG MIN
Public attitude towards activities	+	points	0,047	0,0093 AVG MIN	0,0127 AVG MIN	0,0127 AVG MIN	0,0128 AVG MIN
Change of industrial technology	+	points	0,047	0,0116 AVG MIN	0,0116 AVG MIN	0,0116 AVG MIN	0,0123 AVG MIN
Development of alternative energy sources	+	points	0,034	0,0081 AVG MIN	0,0089 AVG MIN	0,0069 AVG MIN	0,0101 AVG MIN
Environmental regulations	+	points	0,049	0,0121 AVG MIN	0,0109 AVG MIN	0,0095 AVG MIN	0,0144 AVG MIN
Legal regulation of activities	+	points	0,03	0,0076 AVG MIN	0,0072 AVG MIN	0,0094 AVG MIN	0,0060 AVG MIN
Municipal influence	+	points	0,027	0,0004 AVG MIN	0,0057 AVG MIN	0,0070 AVG MIN	0,0051 AVG MIN
Legal basis for infrastr. develop.	+	points	0,023	0,0055 AVG MIN	0,0055 AVG MIN	0,0060 AVG MIN	0,0055 AVG MIN
Profitability degree	+	points	0,065	0,0146 AVG MIN	0,0177 AVG MIN	0,0081 AVG MIN	0,0146 AVG MIN
Competitive environment		points	0,011	0,0024 AVG MIN	0,0021 AVG MIN	0,0030 AVG MIN	0,0030 AVG MIN
Introduction of environment-friendly technologies	+	points	0,034	0,0083 AVG MIN	0,0083 AVG MIN	0,0083 AVG MIN	0,0091 AVG MIN
Dependence on resource provision	-	points	0,013	0,0022 AVG MIN	0,0014 AVG MIN	0,0074 AVG MIN	0,0021 AVG MIN
Integration into internal structructures	+	points	0,019	0,0049 AVG MIN	0,0036 AVG MIN	0,0057 AVG MIN	0,0054 AVG MIN

Dependence on resource provision	-	points	0,013	0,0022 AVG MIN	0,0014 AVG MIN	0,0074 AVG MIN	0,0021 AVG MIN
Integration into internal structures/units	+	points	0,019	0,0042 AVG MIN	0,0036 AVG MIN	0,0057 AVG MIN	0,0054 AVG MIN
Degree of corporate social responsibility	+	points	0,030	0,0092 AVG MIN	0,0096 AVG MIN	0,0090 AVG MIN	0,0090 AVG MIN
Degree of external pollution effect costs		points	0,011	0,0014 AVG MIN	0,0014 AVG MIN	0,0069 AVG MIN	0,0014 AVG MIN
Taxation base level	-	points	0,042	0,0121 AVG MIN	0,0102 AVG MIN	0,0098 AVG MIN	0,0098 AVG MIN
Conditions to fund development projects	+	points	0,034	0,0088 AVG MIN	0,0084 AVG MIN	0,008 AVG MIN	0,0088 AVG MIN
Price of raw materials and energy resources	-	points	0,014	0,0022 AVG MIN	0,0018 AVG MIN	0,0074 AVG MIN	0,0025 AVG MIN
Experience of CEOs	+	points	0,061	0,0152 AVG MIN	0,0149 AVG MIN	0,0152 AVG MIN	0,0106 AVG MIN
Supply of qualified specialists	+	points	0,063	0,0127 AVG MIN	0,0131 AVG MIN	0,0134 AVG MIN	0,0138 AVG MIN
Price of labour resources	-	points	0,05	0,0148 AVG MIN	0,0148 AVG MIN	0,0125 AVG MIN	0,0079 AVG MIN
Readiness to select and use innovations	+	points	0,052	0,0120 AVG MIN	0,0132 AVG MIN	0,0128 AVG MIN	0,0132 AVG MIN
Cooperation with science establishments	+	points	0,027	0,0058 AVG MIN	0,0049 AVG MIN	0,0073 AVG MIN	0,009 AVG MIN
Influence of stakeholder groups		points	0,020	0,007 AVG MIN	0,007 AVG MIN	0,007 AVG MIN	0,007 AVG MIN
Company's income capitalisation value	-	thousand LTL	0,2	0,1051 AVG MIN	0,0084 AVG MIN	0,0849 AVG MIN	0,0017 AVG MIN
Company's rated capacity	+	MW	0,3	0,0952 AVG MIN	0,0107 AVG MIN	0,1904 AVG MIN	0,0087 AVG MIN
Production cost (energy price)	-	ct/kwh	0,5	0,1486 AVG MIN	0,0608 AVG MIN	0,2095 AVG MIN	0,0811 AVG MIN
The sums of weighted normalized maximizing (projects 'pluses') indices of the alternative				0,3040	0,2195	0,3962	0,2100
The sums of weighted normalized minimizing (projects 'minuses') indices of the alternative				0,2908	0,1079	0,3487	0,1169
Significance of the alternative				0,4708	0,5304	0,5157	0,5227
Priority of the alternative				4	1	3	2
Utility degree of the alternative (%)				88,77%	100,01%	97,23%	98,55%

Fig. 7. Compiling the normalised decision-making matrix, determining and comparing the utility degree of objects in the ESIAPVN-DS system

criteria and the sums of maximising and minimising normalised weighted indicators (S_{+j} and S_{-j}), and determines the weight Q_j, priority and utility N_j of the objects.

The system has an important advantage, because it shows the interim results—the analysis of weights of all criteria used in the research, and visual representation of the objects. Criteria weights are analysed by clicking any matrix cell with the value of the selected criterion (links AVG, MN) in the table of multiple criteria analysis of alternatives. The result is a percentage for each selected criterion, which is compared with an equivalent criterion of other objects and shows that an improved value of the criterion may increase the value of the object. The analysis of such results may be an effective and useful tool for comparisons of energy production technologies, for operational analysis, for management decisions and operational improvement plans, and for performance assessments. In the same model window, it is possible to choose an object, to click on the cell with the data of the replacement cost criterion, and the system will show the value of the object revised considering the effect of environmental factors.

The typical input data used in decision support systems for value measurements are the features of the objects and the values of the objects determined using the sales comparison approach. It is hard to come by comparable items when the valuation concerns energy objects and energy production technologies. Moreover, the value of past transactions sometimes does not reflect the real value. Thus the replacement cost of an object (technological facilities) is selected as the initial value in the ESIAPVN-DS system. The

replacement cost of energy objects is generally well above the income generated by the objects. The replacement cost, however, is closer to the value suggested, in legal acts that regulate accounting and valuation, and in European valuation standards, for accounting of assets of economic infrastructure companies (European Valuation Standards, 2009).

To boost the reliability of the utility analysis and market value revisions for energy objects, the main quantitative criterion (replacement cost) is accompanied by additional quantitative criteria—the rated capacity of energy objects (MW), and production cost (cnt/kwh). The system measures the value of the energy objects in question by assessing the values of qualitative and quantitative criteria. This research also includes a quantitative criterion: the income value of the assets of the energy object determined by independent property valuators. Several integrated values (replacement cost and income) make the measured utility, priority and market value more reliable. The ESIAPVN-DS system lets to analyse the object's value using a rather large set of criteria for thorough description of the environment in question. The set may be supplemented or revised, if needed. The reliability of the determined value depends on the importance of selected quantitative and qualitative criteria.

The link "Computer aided development of the feasible alternatives" of the decision support system leads to automatically assessed weights for selected criteria of the objects, and the utility degree of the objects. The best result—the highest utility degree—belongs to Kaunas Hydroelectric Power Plant, followed by the experimental geothermal power plant. The geothermal power plant benefited from its use of renewable resources, and favourable assessment of the environment. Among the advantages of Kaunas Hydroelectric Power Plant are high profitability, positive environmental aspects, favourable public opinion about its activities, and low production costs. Stakeholders also see these objects in a favourable rather than hostile light.

The Lithuanian Power Plant came third by its utility. This power plant is marked by high production costs, negative effect on the environment, high environmental costs, it also depends on increasing prices of raw materials, and the access to resources. But its technical indicators (capacity), as well as the ratio between the produced amounts of energy and the cost of asset generation, make a positive impact. Kruonis Pumped-storage Hydroelectric Power Plant has the lowest utility among all objects. Economic criteria are responsible: rather high cost of produced energy, and high replacement cost. Low volumes of production (electric power) at that time also made impact in the research. Now the power plant is working at full capacity, thus the results would probably be better.

The utility degrees and values of the companies in question measured by the ESIAPVN-DS system are rather logical; also the results come fast, the results and conclusions give more information, the system provides recommendations and prevents errors. The values obtained using the traditional methods and included in the set of quantitative criteria ensure a more reliable result; they enable its comparison, and its use as a basis to validate the measured value. They also help to make a decision on the final value of the object.

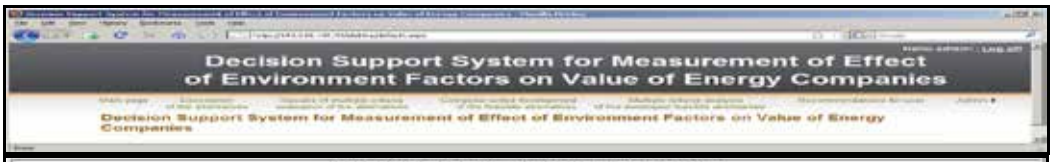
4.3.1 Recommendations in the ESIAPVN-DS System

The problem is how to define an efficient energy production technologies life cycle when a lot of various interested parties are involved, the alternative project versions come to hundreds thousand and the efficiency changes with the alterations in the micro, meso and macro environment conditions and the constituent parts of the process in question. Moreover, the realization of some objectives seems more rational from the economic and

ecological perspectives thought from the other perspectives they have various significance. Therefore, it is considered that the efficiency of energy production technologies life cycle depends on the rationality of its stages as well as on the ability to satisfy the needs of the interested parties and the rational character of the micro, meso and macro environment conditions.

Formalized presentation of the multiple criteria analysis (see Table 8) shows how changes in the micro, meso and macro environment and the extent to which the goals pursued by various interested parties are satisfied cause corresponding changes in the value and utility degree of diferent energy production technologies. With this in mind, it is possible to solve the problem of optimisation concerning satisfaction of the needs at reasonable expenditures. This requires the analysis of energy production technologies versions allowing to find an optimal combination of different interested parties goals pursued, micro, meso and macro environment conditions and finances available. (Sliogeriene et al., 2009).

The ESIAPVN-DS system gives comprehensive information about the quantitative effect on the value by the environmental factors of energy companies: the system analyses the effect of each criterion separately. For example, a click on any selected criteria value in its cell (links AVG, MN) within the matrix of alternatives (menu item "Descriptions of the alternatives") activates automatic assessment and the system offers a recommendation to increase the company's value by changing the criterion respectively (Fig. 8).



Quantitative and qualitative information pertinent to alternatives

Criteria describing the alternatives	*	Measuring units	Weight	- Compared alternatives - Possible improvement of the analysed criterion in % - Possible increase of the market value of the alternative in % through increased value of the aforementioned criterion			
				Kruoto PSPP	Kauno HPP	Lietuvos PP	Geothermal PP
EU regulation of electricity-related activities	+	points	0,049	29 (10,34%)(0,253%)	31 (3,23%)(0,079%)	29 (10,34%)(0,253%)	32 (0%)(0%)
Internal (national) policy towards this sector	+	points	0,023	20 (0%)(0%)	15 (33,33%)(0,383%)	17 (17,65%)(0,203%)	14 (42,86%)(0,492%)
Relations with national authorities	+	points	0,027	17 (11,76%)(0,159%)	17 (11,76%)(0,159%)	19 (0%)(0%)	16 (18,75%)(0,253%)
Trends of economic change	+	points	0,028	24 (8,33%)(0,116%)	18 (44,44%)(0,621%)	26 (0%)(0%)	17 (52,94%)(0,74%)
Changes of consumption	+	points	0,03	22 (18,18%)(0,272%)	19 (36,84%)(0,552%)	26 (0%)(0%)	18 (44,44%)(0,666%)
Investment conditions	+	points	0,047	34 (0%)(0%)	30 (13,33%)(0,313%)	29 (17,24%)(0,405%)	20 (70%)(1,643%)
Public attitude towards activities	+	points	0,047	22 (36,36%)(0,853%)	30 (0%)(0%)	30 (0%)(0%)	29 (3,45%)(0,081%)
Change of industrial technology	+	points	0,047	29 (6,9%)(0,162%)	29 (6,9%)(0,162%)	29 (6,9%)(0,162%)	31 (0%)(0%)
Development of alternative energy sources	+	points	0,034	20 (25%)(0,424%)	22 (13,64%)(0,231%)	17 (47,06%)(0,799%)	25 (0%)(0%)
Environmental regulations	+	points	0,049	27 (18,52%)(0,453%)	31 (3,23%)(0,079%)	19 (68,42%)(1,674%)	32 (0%)(0%)
Legal regulation of activities	+	points	0,03	20 (10%)(0,15%)	19 (15,79%)(0,236%)	22 (0%)(0%)	18 (22,22%)(0,333%)
Municipal influence	+	points	0,027	25 (0%)(0%)	17 (47,06%)(0,634%)	23 (8,7%)(0,117%)	15 (66,67%)(0,800%)
Legal basis for infrastr. develop.	+	points	0,023	15 (20%)(0,23%)	15 (20%)(0,23%)	18 (0%)(0%)	15 (20%)(0,23%)
Profitability degree	+	points	0,055	29 (20,69%)(0,568%)	35 (0%)(0%)	16 (118,75%)(3,261%)	29 (20,69%)(0,568%)
Competitive environment	-	points	0,011	8 (12,5%)(0,069%)	7 (0%)(0%)	11 (36,36%)(0,2%)	11 (36,36%)(0,2%)
Introduction of environment-friendly technologies	+	points	0,034	22 (9,09%)(0,154%)	22 (9,09%)(0,154%)	22 (9,09%)(0,154%)	24 (0%)(0%)
Dependence on resource provision	-	points	0,013	13 (38,46%)(0,25%)	8 (0%)(0%)	43 (81,4%)(0,528%)	12 (33,33%)(0,216%)
Integration into internal structures/oras	+	points	0,019	14 (35,71%)(0,339%)	12 (58,33%)(0,553%)	19 (0%)(0%)	18 (5,56%)(0,053%)

Fig. 8. Calculations in the matrix of alternatives revealing how the criteria affect the value

The system's user may select any criterion relevant to the analysis in question. The ESIAPVN-DS system helps to get automated and unbiased recommendations about the effect of individual criteria. The recommendations are grounded on the values and weights of specific criteria, as well as on comparisons of the objects. The ESIAPVN-DS system lets to analyse individual criteria: not only to find the criteria with the biggest impact on the value, but also to compute the chance to improve the criteria. It is then possible to make reasonable decisions and choose for companies the development trends, innovations, optimisation of economic indicators, reduction of hostility among stakeholder groups, or initiation of amendments in the legislative basis. Figure 9 shows some recommendations suggested by the ESIAPVN-DS system.

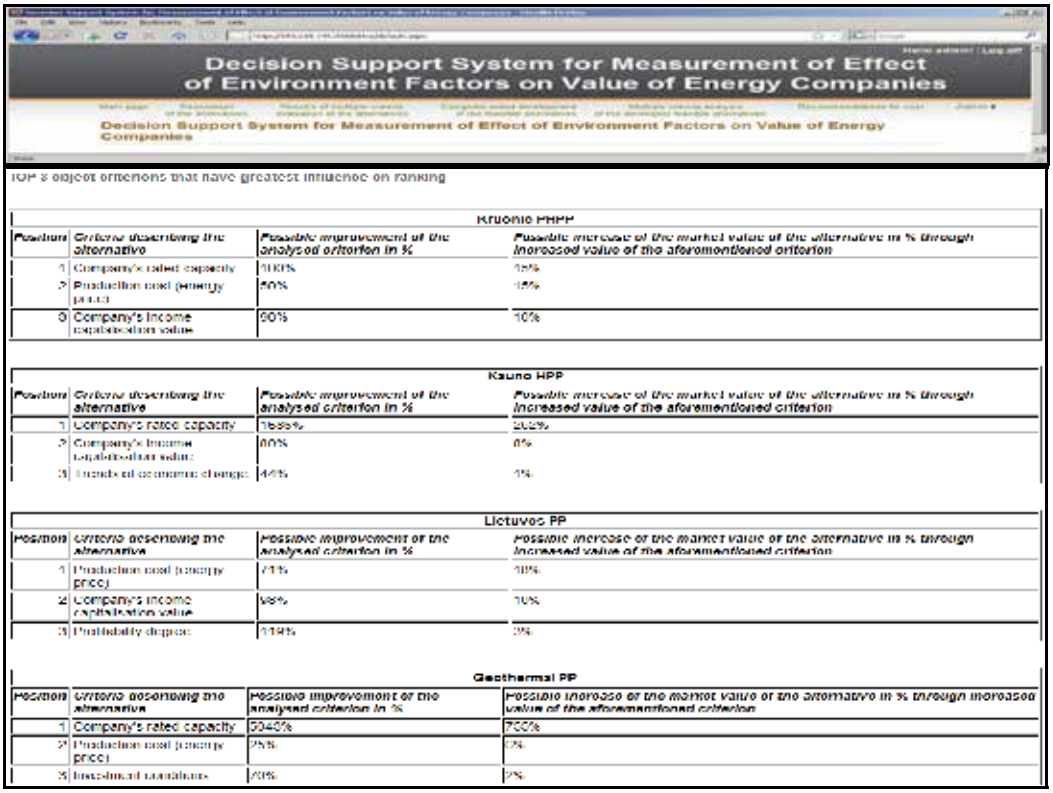


Fig. 9. Recommendations in the ESIAPVN-DS system

We shall use the criterion "legal regulation of activities" of the experimental geothermal power plant to illustrate the potential of the recommender module. Experts gave 18 points to this criterion of the experimental geothermal power plant (the worst legal regulation of activities among all energy objects in question). Calculations show that the legal regulation of activities may be improved by about 22 %. Such improvement of the legal regulation of activities by 22 % would raise the market value of the geothermal power plant by 2 %. The European Union strongly supports the expansion of alternative energy and encourages the governments of member states to reorganise their legislative basis respectively, thus creating favourable conditions for the expansion of alternative energy production sources.

This criterion may definitely be improved and the improvement may be achieved by efforts of CEOs and politicians.

The developed ESIAPVN-DS system ensures informative and unbiased results, which may be used in assessment of diverse energy objects aiming to determine their utility and the efficiency of their environment, but in other types of analysis as well. This system is experimental and not used for practical applications, because it needs further improvement and enhancement of individual functions, as well as continuity of studies. The current version of the system, however, may be used as an extra method or tool to control the reliability of results in a range of studies or analyses. The system, after further improvements, may be a useful tool in analyses dealing with the impact of environmental factors on the value, utility, and performance of objects, in identification of the vital values of environmental factors, and as an aid to plan the potential activities, to choose development scenarios, and to assess technologies. The ESIAPVN-DS system uses the set of criteria describing the environmental factors and the criteria weights and thus may analyse a variety of companies operating in the sector, may accumulate the data and use it later in analyses of other objects. In analysis of the development trends of the energy sector, in selection of alternatives, or in measuring of the utility degrees of individual energy production technologies, the main research interest is selection of the criteria and building of a system of universal criteria for the sector's analysis. Important aspects are choice of adequate criteria significant in the sector, finding the relation between criteria of various levels, and assessment of their weights. Then it would be possible to make the decision support systems for the analysis of the sector more universal, and to expand their application.

5. Conclusions

In the energy sector, decisions encompass interrelated solutions at various levels: selection of development scenarios in the energy sector; decisions on implementation of political instruments; or decisions on the choice of specific technologies for energy production in the future, and on promotion and implementation of specific energy projects. The analysis of alternative decisions made at different levels demands for relevant criteria for the assessment of such alternatives, as well as for indicators to be used in quantitative and qualitative evaluation of the alternatives. The methods based on multiple criteria analysis are promising when one has to analyse the energy sector and energy production technologies. Moreover, these methods have advantages over the traditional analysis methods based on economic data. The multiple criteria methods have tools that help to consider the full set of environmental factors of the object, and to integrate significant economic indicators. They also eliminate the bias of analysts and ground the assumptions on a comprehensive market analysis.

The experts assessed the impact on the value by environmental factors and their assessment suggests that the experience of CEOs, readiness to introduce innovations, environmental factors, and corporate social responsibility are important factors that affect the activities of the selected objects operating in the sector in question. The effect of these factors on operating decisions and the value of objects may only be assessed using innovative methods based on mathematical analysis, for example, multiple criteria complex proportional method for measurement of the utility degree and the market value.

The experimental Decision Support System for Measurement of the Effect by the Environmental Factors on the Value of Energy Companies (ESIAPVN-DS), based on the algorithm suggested by the authors and on multiple criteria analysis, enables a more comprehensive process for solution framing and has the following advantages:

- the system helps to quickly measure the utility, priority, and value of complex objects using contemporary methods, it helps to consider and analyse substantially more environmental factors, and to make integrated assessments of quantitative and qualitative indicators describing the objects in question;
- the DSS measures the value based on a comprehensive analysis of the environment, and grounds the assumptions with better accuracy; the measuring process not only shows the final value, but also offers comprehensive interim results helpful in decision-making at various levels (recommendations about the effect of separate environmental factors on the value, and about the possibilities to mitigate the negative effect of the factors thus improving the operating environment);
- the system can be easily supplemented, improved and then used to frame various solutions: to assess the impact of environmental factors, to identify the vital values of factors, to plan the courses of action, to submit reliable information to various institutions and, finally, to analyse the reasons behind changing value.

6. References

Almeida, A.T. (2005). Multicriteria modelling of repair contract based on utility and ELECTRE I method with dependability and service quality criteria. *Annals of Operations Research*, Vol.138, No.1, (September 2005), p.p. 113–126, ISSN 1572-9338

Ashina, S.& Nakata, T. (2008). Quantitative analysis of energy-efficiency strategy on CO_2 emissions in the residential sector in Japan – Case study of Iwate prefecture. *Applied Energy,* Vol.85, No.4, (April 2008), p.p. 204–217, ISSN 0306-2619

Benavides, J. A. C. & Prado J. C. (2002). Creating an Expert System for Detailed Scheduling. *International Journal of Operations & Production Management*, Vol.22, No.7, p.p. 806-819, ISSN 0144-3577

Bergey, P. K.; Ragsdale, C. T. & Hoskote, M. (2003). A decision support system for the electrical power districting problem. *Decision Support Systems*, Vol.36, No.1, (September 2003), p.p. 1-17, ISSN 0167-9236

Bertin, J. (1983). *Semiology of Graphics*. Madison, WI: University of Wisconsin Press, ISBN 0299090604

Booty, W. & Wong, I. (2010). Case Studies of Canadian Environmental Decision Support Systems. *Decision Support Systems,* INTECH, Olajnica 19/2, 32000 Vukovar, Croatia, ISBN 978-953-7619-64-0, Available from <www.sciyo.com>

Brans. J.P. & Mareschal, B. (2005). PROMETHEE methods, in: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Chapter 5. *Multiple CriteriaDecision Analysis: State of the Art Surveys*. Springer, , p.p. 163–195, ISBN 978-0-387-23067-2

Brito, A. J. & Almeida, A. T. (2008). Multi-attribute risk assessment for risk ranking of natural gas pipelines, *Reliability Engineering & System Safety,* Vol.94, No.2, (February 2009), p.p. 187-198, ISSN 0951-8320

Commision of the Europian Communities. Green Paper. A European Strategy for Sustainable, Competitive and Secure Energy. KOM (2006) 105 final, Available from: <http://europa.eu/documentation/official-docs/green-papers/index_en.htm>

Chen, S. J., & Hwang, C. L. (1992). Fuzzy Multiple Attribute Decision Making Methods and Applications. *Lecture Notes in Economics and Mathematical Systems*, 375, Springer – Verlag, Berlin, Germany.

Deok, J.L. & Jooho, H. (2008). Decision support for selecting exportable nuclear technology using the analytic hierarchy process: A Korean case. *Energy Policy*, Vol.38, No.1, (January 2010), p.p. 161-167, ISSN 0301-4215

Drozdz, M. (2003). Construction of energy and environmental information systems. *Applied Energy* , Vol.76, No.1-3, (September 2003), p.p. 279–288, ISSN 0306-2619

Europiann Valuation Standards (EVS), 2009.(The Blue Book), Available from: <http://www.tegova.org/en/p4912ae3909e49#p4912ae5f0a65e>

Figueira, J.; Greco, S. & Ehrogott, M. (Eds.). (2005). Multiple Criteria Decision Analysis: State of the Art Survey. Springer, Berlin, Vol.78, ISBN: 978-0-387-23067-2

Guitouni, A. & Martel, J.M. (1998). Tentative guidelines to help choosing an appropriate MCDA method. *European Journal of Operational Research*, Vol.109, No.2. (1 September 1998), p.p. 501–521, ISSN 0377-2217

Higgins, A. J.; Hajkowicz, S.; & Bui, E. (2008). A multi-objective model for environmental investment decision making. *Computers and Operations Research*, Vol.35, No.1, (January 2008), p.p. 253–266, ISSN 0305-0548

Hwang, C.L. & Yoon, K. (1981). *Multiple Attribute Decision Making. Methods and Applications*. Berlin, Heidelberg, New York: Springer – Verlag

Hwang, M. & Quigley, J. M. (2004). Selectivity, quality adjustment and mean reversion in the measurement of huse values, *The Journal of Real Estate Finance and Economics*, Springer, Vol.28, No.2-3, (March 2004) p.p. 161–178, ISSN 0895-5638

Jeroen C.J.M. van den Bergh. (1999). *Handbook of Environmental and Resource Economics*. Cheltenham, UK ; Northampton, Mass. : Edward Elgar Pub

Kaklauskas, A. (1999). Multiple criteria decision support of building life cycle: Research report presented for habilitation. Vilnius Gediminas Technical University, Vilnius

Kendall, M. G. (1970). Rank correlation methods, 4th ed, London: Griffin.

Khuri, A. I. (2002). Graphical evaluation of the adequacy of the method of unweighted means. *Journal of Applied Statistics*, Vol.29, No.8, (November 2002), p.p 1107-1119, ISSN 0266-4763

Lu, M. & So, R. W. (2005). Return relationships between listed banks and real estate firms: evidence from seven asian economies. The Journal of Real Estate Finance and Economics, Vol.31, No. 2, p.p. 189–206, ISSN 0895-5638

Martinson, M. G. & Davison. R. M. (2007). Strategic decision making and support systems: Comparing American, Japanese and Chinese management. *Decision Support Systems*. Vol.43, No.1, (February, 2007), p.p. 284-300, ISSN 0167-9236

Masood, S. H. & Soo, A. A. (2002). Rule Based Expert System for Rapid Prototyping System Selection. *Robotics & Computer–Integrated Manufacturing*, Vol.12, No. 3-4, (June-August 2002), p.p. 267–275, ISSN 0736-5845

Morais, D.C. & Almeida, A. T. (2007). Group decision-making for leakage management strategy of water distribution network. *Resources, Conservation and Recyclin,* Vol.52, No.2, (December 2007), p.p. 441–459, ISSN 0921-3449

Neves, L. P.; Martins, A. G.; Antunes, C. H. & Dias, L. C. (2008). A multi-criteria decision approach to sorting actions for promoting energy efficiency. *Energy Policy*. Vol. 36, No.7, (July 2008), p.p. 2351-2363, ISSN 0301-4215

Patlitzianas, K.D. & Psarras, J. (2007). Formulating a modern energy companies' environment in the EU accession member states through a decision support methodology. *Energy Policy*, Vol.35, No.4, (April 2007), p.p. 2231-2238, ISSN 0301-4215

Pillet, G.; Maradan, D.; Mayor, K.; Stephani, E.; & Zein, K. (2005). Mezo environment–economic analyses. Methodology and main results – industry and urban communities in Arab countries. International Congress Sustainable Management in Action, SMIA, 2 Congrès international. Geneva, 19-20 September, 2005.

Podvezko, V. (2005). Agreement of expert estimates. *Technological and Economic Development of Economy*, Vol.11, No.2, p.p. 101–107, ISSN 2029-4913

Podvezko, V. & Podviezko, A. (2010). Dependece of Multi-Criteria Evaluatio Result on Choice of Preference Functions and Their Parameters. *Technological and Economic Development of Economy,* Vol.16, No.1, p.p. 143-158, ISSN 2029-4913

Saaty, T. L. (2000). *Fundamentals of the Analytic Hierarchy Process*. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15413.

Scholz, R.W. & Tietje, O. (2002). *Embedded Case Study Methods. Integrating Quantitative and Qualitative Knowledge*. London: Sage Publications Inc., ISBN 0-7619-1946-5

Sims, S.; Dent, P.; & Oskrochi, R. (2008). Moddeling the impact of winds farms on house prices in the UK, *International Journal of Strategic property Management,* Vol.12, No.4, p.p. 251–269, (November 2008), ISSN 1648-715X

Shen, Y-C.; Lin, G. T. R.; Li, K-P. & Yuan, B. J. C. (2010). An assessment of exploiting renewable energy sources with concerns of policy and technology. *Energy Policy*. Vol.38, No.8, (August 2010), p.p. 4604-4616, ISSN 0301-4215

Shilei, Lv. & Yong, Wu. (2009). Target-oriented obstacle analysis by PESTEL modeling of energy efficiency retrofit for existing residential buildings in China's northern heating region. *Energy Policy*. Vol.37, No.6, (June 2009), p.p. 2098-2101, ISSN 0301-4215

Sliogeriene, J.; Kaklauskas, A.; Zavadskas, E.K.; Bivainis, J.& Seniut, M. (2009). Environmet factors of energy companies and their effect on value: analysis model and applied methods, *Technological and Economic Development of Economy*, Vol.15, No.3, p.p.490–521, ISSN 2029-4913

Staniskis, J. K. & Stasiskiene, Z. (2006). Environmental management accounting in Lithuania: exploratory study of current practices, opportunities and strategic intent. *Journal of Cleaner Production,* Vo.14, No.14, p.p. 1252–1261, ISSN 0959-6526

Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Application*, Wiley Series in Probability and Mathematical Statistics, John Wiley, New York, 546 pp., ISBN: 047188846X

Streimikiene D. & Mikalauskiene A. (2009). Application of Integrated Indicators for Monitoring the National Energy Strategy. *Power engineering*. Vol.55, No. 3, p.p. 158–166, ISSN 0235-7208

Streimikiene, D. & Pusinaite, R. (2008). External cost of electricity generation in Lithuania. *Environment Research, Engineering and Management,* Vol.2, No.44, p.p. 34-40, ISSN 1392-1649

Sueyoshi, T.& Tadiparthi, G. R. (2008). An agent-based decision support system for wholesale electricity market. *Decision Support Systems*, Vol.44, No.2, (January 2008), p.p. 425-446, ISSN 0167-9236

Svensson J. S. (2002). The Use of Legal Expert Systems in Administrative Decision Making. In: Ake Grönlund (ed.). Chapter 8. *Electronic Government: Design, Applications and Management*.), Hershey, London (etc.), Idea Group Publishing, ISBH 1930708-19-X

Tam, C. M.; Tong, T. K. L. & Chiu, G. W. C. (2006). Comparing non-structural fuzzy decision support system and analytical hierarchy process in decision-making for construction problems. *European Journal of Operational Research*, Vol.174, No.2, (16 October 2006), p.p. 1317–1324, ISSN: 0377-2217

Triantaphyllou, E. (2000). *Multi–criteria Decision Making Methods: a Comparative Study. Dordrecht*. The Netherlands: Kluwer Academic Publishers (now Springer), ISBN 0792366077

Yue, Cheng-Dar & Grant, Gwo-Liang Y. (2007). Decision support system for exploiting local renewable energy sources: A case study of the Chigu area of southwestern Taiwan. *Energy Policy*, Vol.35, No.1, (January 2007), p.p. 383-394, ISSN 0301-4215

Zavadskas, E. K.; Peldschus, F. & Kaklauskas, A. (1994). *Multiple criteria evaluation of projects in construction*. Vilnius, Technika. 1994, ISBN 9986-05-046-4

Zavadskas, E. K. & Kaklauskas, A. (2008). Model for Lithuanian construction industry development. *Transformations in Business & Economics*, Vol.7, No.1(13), p.p. 152–168, ISSN 1648 - 4460

Zeng, J.; Min, A. & Smith, N. J. (2007). Application of a fuzzy based decision making methodology to construction project risk assessment. *International Journal of Project Management.* Vol.25, No.6, (August 2007), p.p. 589–600, ISSN 0263-7863

Zimmermann, H. J.& Gutsche, L. (1991). *Multi-Criteria Analyse*. Berlin/Heidelberg, Germany:Springer-Verlag, 1991

Von Winterfeldt, D. & Edwards, W. (1986). *Decision Analysis and Behavioural Research*. Cambridge University Press, Cambridge, 1986

Edited by Chiang Jao

This series is directed to diverse managerial professionals who are leading the transformation of individual domains by using expert information and domain knowledge to drive decision support systems (DSSs). The series offers a broad range of subjects addressed in specific areas such as health care, business management, banking, agriculture, environmental improvement, natural resource and spatial management, aviation administration, and hybrid applications of information technology aimed to interdisciplinary issues. This book series is composed of three volumes: Volume 1 consists of general concepts and methodology of DSSs; Volume 2 consists of applications of DSSs in the biomedical domain; Volume 3 consists of hybrid applications of DSSs in multidisciplinary domains. The book is shaped upon decision support strategies in the new infrastructure that assists the readers in full use of the creative technology to manipulate input data and to transform information into useful decisions for decision makers.

ISBN 978-953-51-5534-8
9 789535 155348

Photo by AndreasG / Shutterstock

IntechOpen